ESD-TR-66-671, Vol. II

PHASE II FINAL REPORT ON USE OF AIR FORCE ADP EXPERIENCE TO ASSIST AIR FORCE ADP MANAGEMENT

VOLUME II

PHASE II ACTIVITIES

Alan J. Gradwohl
George S. Beckwith
Stanton H. Wong

December 1966

TACTICAL PLANNING DIVISION
DIRECTORATE OF PLANNING AND TECHNOLOGY
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts

Distribution of this document is unlimited.

ESD-TR-66-671, Vol. II

PHASE II FINAL REPORT ON USE OF AIR FORCE ADP EXPERIENCE TO ASSIST AIR FORCE ADP MANAGEMENT

VOLUME II

PHASE II ACTIVITIES

Alan J. Gradwohl
George S. Beckwith
Stanton H. Wong

December 1966

TACTICAL PLANNING DIVISION
DIRECTORATE OF PLANNING AND TECHNOLOGY
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts

FOREWORD

This is Volume II of a three-volume final report prepared by the Information Systems Division of Planning Research Corporation, Los Angeles, California, under contract number AF 19(628)-5988, project number 7990. The Air Force Project Officer was Major George H. Montague, Electronic Systems Division, ESLT. Work on the project was performed under the direction of Alan J. Gradwohl, PRC Project Manager, from 16 February 1966 to 15 December 1966, with Stanton H. Wong in charge of producing the final report. Contributors to this volume of the final report from the PRC staff include:

Lynn R. Ostrick
Lawrence G. Wimpey
Wolford O. Wootan, Jr.

This technical report has been reviewed and is approved.

George H Montague
GEORGE H. MONTAGUE
Major, USAF
Project Officer

Charles G. Johnson
CHARLES G. JOHNSON
Colonel, USAF
Chief, Tactical Planning Division
Directorate of Planning and Technology

FOREWORD

This is Volume II of a three-volume final report prepared by the Information Systems Division of Planning Research Corporation, Los Angeles, California, under contract number AF 19(628)-5988, project number 7990. The Air Force Project Officer was Major George H. Montague, Electronic Systems Division, ESLT. Work on the project was performed under the direction of Alan J. Gradwohl, PRC Project Manager, from 16 February 1966 to 15 December 1966, with Stanton H. Wong in charge of producing the final report. Contributors to this volume of the final report from the PRC staff include:

Lynn R. Ostrick
Lawrence G. Wimpey
Wolford O. Wootan, Jr.

This technical report has been reviewed and is approved.

George H Montague
GEORGE H. MONTAGUE
Major, USAF
Project Officer

Charles G. Johnson
CHARLES G. JOHNSON
Colonel, USAF
Chief, Tactical Planning Division
Directorate of Planning and Technology

ABSTRACT

This is Volume II of a three-volume final report that covers Phase II of a three-phase project on the Use of Air Force ADP Experience to Assist Air Force ADP Management. In Phase I, a feasible concept and preliminary approach to using experience was synthesized; in Phase II, the approach was refined, the concept was validated, and the potential use of experience was broadened; and in Phase III, the improved and expanded approach will be implemented Air Force-wide.

Volume I of the final report covers the following: the history of the project; conclusions of Phase II and recommendations for Phase III; and summaries of Phase II activities, the Phase III concept and plan, and the pilot version of the ADP Experience Handbook and Primer. Volume II reviews the four major activities of Phase II: data collection, data analysis, ADP Experience Handbook development, and Phase III planning. Volume III presents the detailed Phase III operational concept and development plan, followed by a summary of cost and benefits.

This is Volume II, in which the four major activities of Phase II are described. The design of the data collection questionnaire was based on the ADPS model (a concept of a "total" ADPS) and the workload model representing attributes of an ADPS. Data were collected on a stratified 18-ADPS sample, and the statistical analysis of these data produced five cost estimation equations. In addition, the data were used to produce a seven-page system description of each ADPS, which became the core of the ADP Experience Handbook. A Phase III operational concept and development plan was also synthesized.

TABLE OF CONTENTS
(Continued)

LIST OF FIGURES

LIST OF TABLES

I. INTRODUCTION

This is Volume II of a three-volume final report that marks the completion by Planning Research Corporation of a research study on the Use of Air Force ADP Experience to Assist Air Force ADP Management. The study is the second phase of a three-phase project; Phase II is to validate and refine concepts developed in Phase I and to develop an operational concept and plan for implementation in Phase III.

The purpose of the final report is to present the objectives, activities, findings, and conclusions of Phase II and to submit an operational concept and development plan for Phase III. These are reported in Volume II and Volume III, respectively. In addition, the pilot version of the ADP Experience Handbook and a Primer that serves as an elementary text for training potential users of the handbook are produced as two separate volumes distinct from this final report (refer to PRC documents R-930 and R-931). Volume I provides a concise summary of Volumes II and III, and a brief description of the ADP Experience Handbook and Primer.

The purpose of Volume II is to present the objectives, activities, findings, and conclusions of Phase II in detail. This volume is directed to those audiences that desire a complete description of all or any part of the activities in Phase II. Volume I provides a concise summary of Volumes II and III and a brief description of the ADP Experience Handbook and Primer.

This volume is organized into four major sections covering the major activities of Phase II: data collection, data analysis, experience handbook development, and Phase III planning. The section on data collection covers model development, ADPS sample, data collection, and data reduction. The section on data analysis deals with refinement of the workload model, testing for subpopulations, and derivation of the cost estimation equations. The section on experience handbook development reviews the development of cost estimation graphs, system descriptions, development of indexes, and construction of the Primer. The section on Phase III planning discusses the development of the Phase III operational concept and plan. Seven appendixes support the text with data and procedures.

To aid the reader, especially for the data collection and data analysis sections, a brief classification and definition of terminology associated with models and variables as used in this volume will be given here. The relationships among the various terms are shown in Figure 1.

The dependent variables are classified into planning factors, and the independent variables are classified into estimating factors. The dependent variables are to be estimated by the independent variables, which are

then called predictors. The independent variables also are referred to as workload descriptors and explanatory variables. The workload model consists of the entire set of variables, both dependent and independent. The cost model is a subset of the workload model, consisting of all the cost variables that comprise the total cost of an ADPS and the workload descriptors that are causally related to each of the cost variables. The regression model is a subset of the cost model that is used in regression analysis; the intercorrelated workload descriptors for each cost variable in the cost model has been removed.

The ADPS model is the concept of the total ADPS and is used as a basis for data collection and system description. The terms "macro-description," "total description," and "system description" are used synonymously when referring to an ADPS.

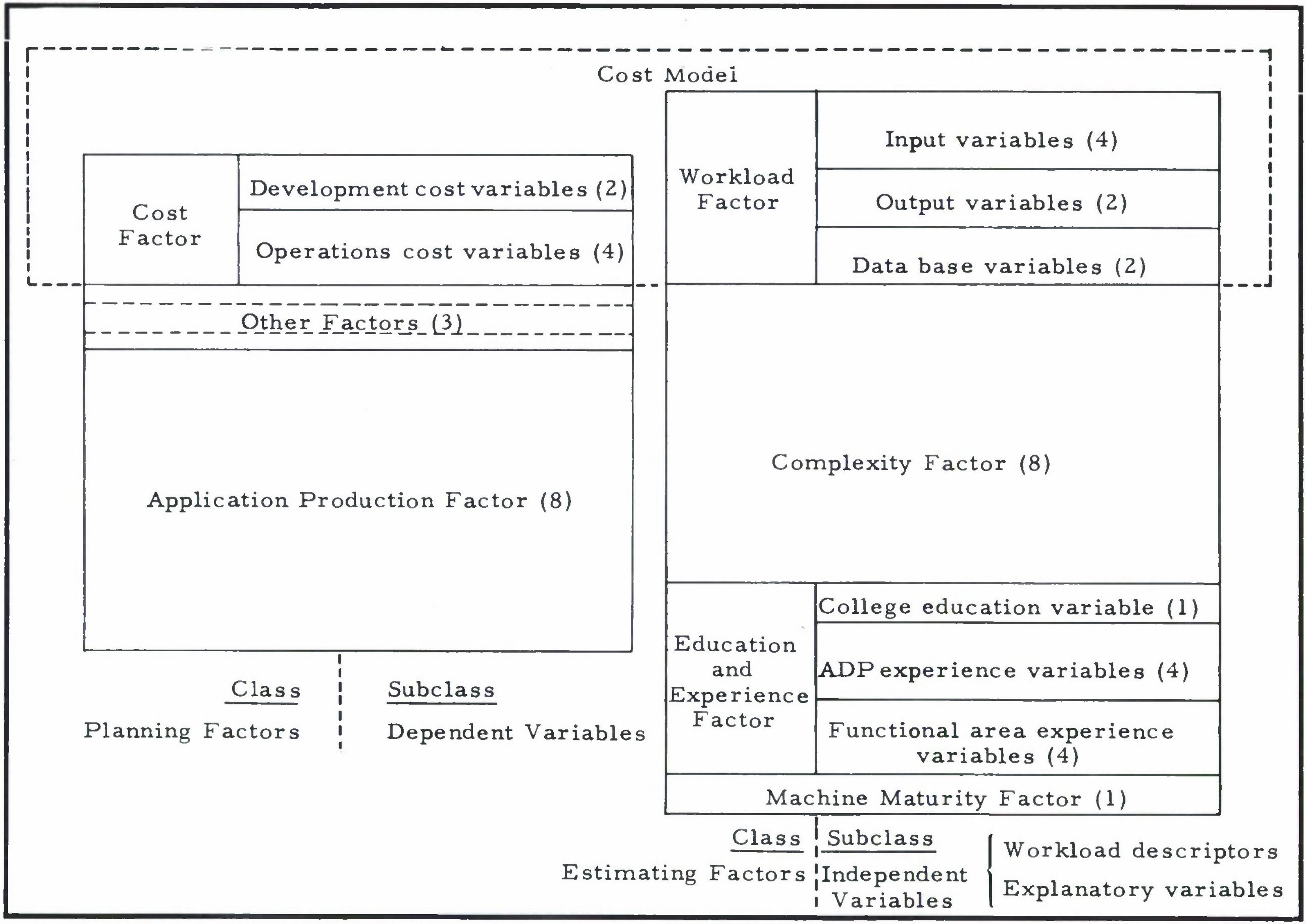


FIGURE 1 - WORKLOAD MODEL

II. DATA COLLECTION

The purpose of this section is to review the data collection activities of Phase II.

A. Objectives

The objectives of data collection for Phase II were twofold:

- To collect sufficient descriptive information to permit macrodescriptions of selected Air Force ADPS.
- To collect sufficient numerical cost data and workload descriptor parameters to permit the development of cost estimating relationships for ADP systems.

To achieve these objectives, the following tasks were performed:

- Development of an ADPS model
- Redefinition of the workload model
- Development of operational definitions and measures for variables
- Restructure of ADPS sample
- Development of data collection procedures
- Reduction of collected data

Each of these tasks is described in the subsequent sections. Finally, a section covering findings completes the section on data collection.

B. ADPS Model

In order to collect sufficient descriptive information to produce a macrodescription of an ADPS, a model representing the concept of a total ADPS was developed. This model served as a basis for and guided the development of the following major activities in data collection:

- Design of the questionnaire
- Collection of data during trips
- Compilation and reduction of first-level data
- Preparation of system summaries for the midpoint report

- Preparation of system descriptions for the pilot version of the ADP Experience Handbook

The principal objective in the development of the ADPS model was to create a concept with the following characteristics:

- Logical breakdown for organization of the interviewing activity
- Ease of explanation to and understanding by the interviewee of the concept
- Organization along lines of information availability
- Compatibility with many forms of Air Force ADP systems

The concept developed for describing the total ADPS was based on the evolution of activities of the ADPS over time. The time axis for the ADPS was divided into four major periods. These were called Proposal Phase, Development Phase, Operations Phase, and Future Plans. These phases were not always clearcut, but, for the purposes of the Phase II study only, they were defined as follows:

- Proposal Phase: This covers the period from the conception of the system to the time the proposal for the system was approved
- Development Phase: This covers the period from the approval of the proposal or the beginning of system design to the time when the system was declared operational
- Operations Phase: This covers the period from the time the systems was declared operational to the present time
- Future Plans: This covers the period beyond the present time

See Table 1 for a schematic representation of the total ADPS concept. Within each phase, the types of data of major interest are itemized.

C. Workload Model

A workload model was defined in Phase I. It was hypothesized that the key to retrieving experience information was workload--quantitative measures of the information processed. The reasons for using workload rather than some other factors were as follows:

- Workload is a direct causal factor for cost and development time

- o Workload is amenable to quantitative measurement
- o Workload should be available in a proposal for an ADPS

Forty numerical workload descriptors were advanced in Phase I as those satisfying the three criteria. These workload descriptors were to be analyzed and evaluated during Phase II by statistical techniques on sampled ADP systems (1) to determine relationships between ADPS workload descriptors and ADPS cost and development time, and (2) to refine those relationships to a well-defined, sensitive, and small set of workload descriptors.

During the initial stages of Phase II, the workload model served two major functions. Firstly, the design of the data collection questionnaire was based on the ADPS model and the workload model. And secondly, the relevant causal factors for use in the regression analysis to derive cost estimating relationships were obtained from the workload model.

During the design of the questionnaire, the original workload model was modified and expanded. Subsequent to data collection, this model was further refined, some variables were dropped, and others were combined. (See subsection II.D.) The modifications and refinements were necessitated by the unavailability of data for some variables. The resulting workload model is schematically depicted in Figure 2, and its function is described below.

The workload model became the basis for development of the cost model. The cost model is comprised of a set of dependent variables called cost factors, which together represented the total cost for development and operations of an ADPS. (See Table 4.) The development of the cost model is described in subsection III.B.6.

The usefulness of the workload model for deriving cost estimating relationships depended entirely on availability of historical data for those sets of variables that represent the characteristics, functions, and costs of the sampled ADPS. The set of workload descriptors are the independent variables or estimating factors in the cost model, and will later be used to derive the cost estimating relationships by a statistical technique called regression analysis.

The workload factor is that set of independent variables that relate to the inputs, outputs, and data base functions of the workload model. The complexity factor relates to the processing functions, the education and experience factor relates to personnel, and the machine maturity factor relates to equipment. Each of the independent variables is also referred to in this report as a <u>workload descriptor</u>.

D. <u>Operational Definitions and Measures of Workload Descriptors</u>

The value of data analysis is heavily dependent on the accuracy and uniformity of the data collected. The accuracy and uniformity of

TABLE 1 - ADPS MODEL FOR PHASE II DATA COLLECTION--
CONCEPT OF TOTAL ADPS

General Description

Proposal	Development	Operations	Future Plans
Content	Technical approach	Organization	
Preparation	Management approach	Manpower	
	Schedule	Workload	
	Manpower	Scheduling	
	Hardware	Utilization	
	Software supplied by others	Program maintenance	
	System design	Hardware	
	Programming	Support programming	
	File conversion	Facility	

Time Axis

System Conceived	Proposal Approved or System Design Begins	Declared Operational	Present Time

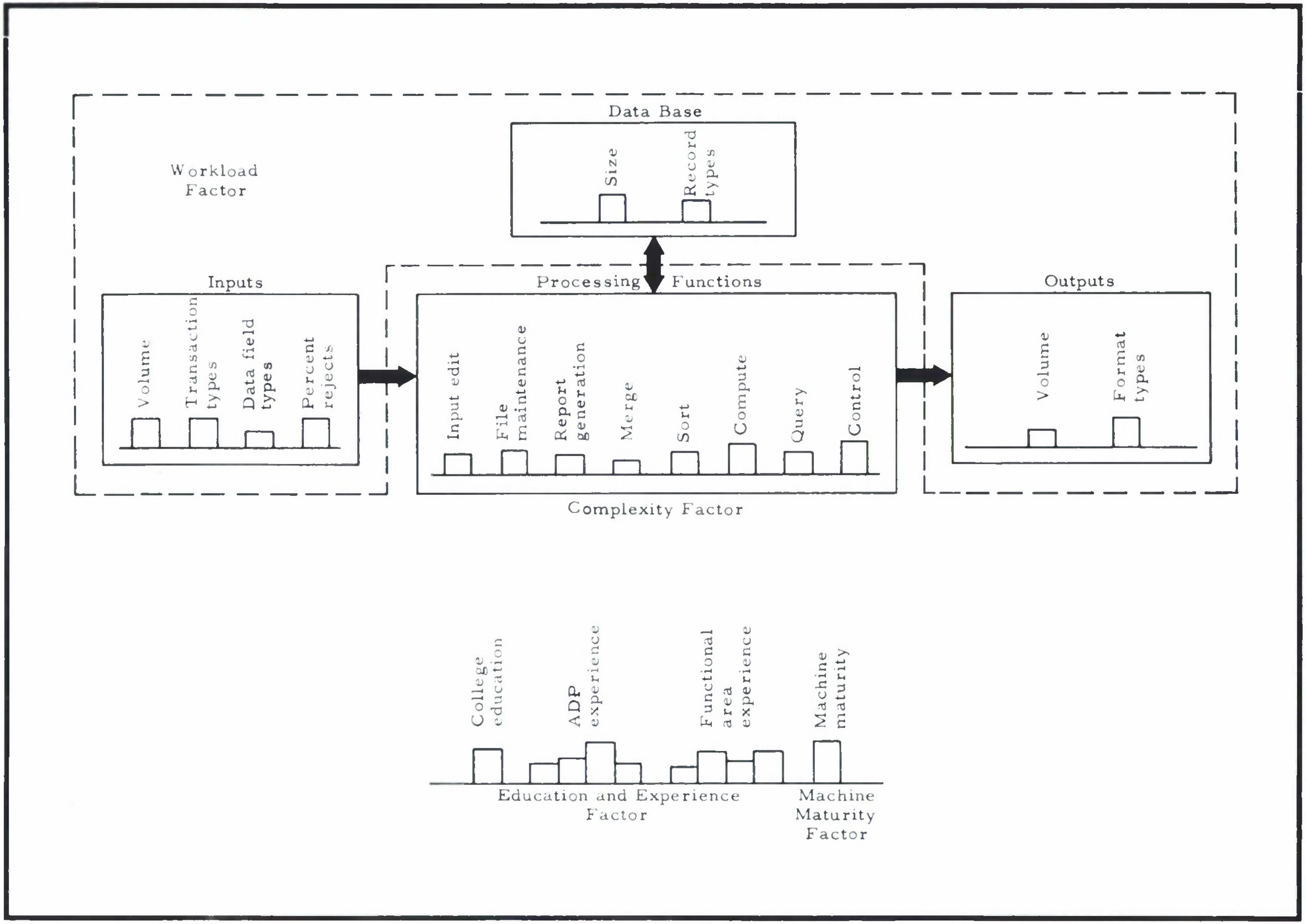


FIGURE 2 - WORKLOAD MODEL (INDEPENDENT VARIABLES--ESTIMATING FACTORS)

the observations, in turn, are dependent to a large degree on the preciseness of definition of the variables in the statistical analysis.

During Phase I, a preliminary workload model was described and workload descriptors defined. At the beginning of Phase II, the definitions of these descriptors were reviewed and redefined during the process of constructing the questionnaire. Following the first data collection trip (see subsection II.F.4), the experience gained was used to modify and enhance the earlier definitions. In addition, several variables were dropped and a number of variables were added. These were categorized into 57 independent variables (estimating factors) and 30 dependent variables (planning factors). See Numerical Data Summary Sheet in Appendix B.

The second and third data collection trips provided additional experience on the form of data that were collectible. The wide variety of systems encountered helped to shake down and test the adequacy of definition of variables. From this experience, the definition of each variable was scrutinized and redefined when necessary. The final list of variables used in the data analysis and their operational definitions can be found in Appendix C. These include 26 independent variables, which constitute the preliminary set of estimating factors, and 7 dependent variables, which constitute the preliminary set of planning factors. The discussion of data reduction (subsection II.G) will describe how this set was obtained.

E. ADPS Sample

On 24 February 1966, project personnel discussed with AFADA and ESD personnel the criteria for selection of the 18 ADP systems to be surveyed. The criteria were those stated in the Phase I report:

- Selected systems must be stratified by size (small, medium, large) and by functional area (similar and dissimilar)
- Selected systems must have undergone a fairly recent development so that data from that phase will still be available
- Selected systems must not present any unusual security problems

On 7 March 1966, AFADA selected the sample of 18 ADP systems; but four of these systems subsequently had to be replaced because further investigation revealed an extreme scarcity of data available for the development phase. The following systems were replaced during data collection:

Original	Replacement
Tech Order Distribution Tinker AFB	Data Services Workload Control Kelly AFB
Inventory Management, Stock Control, Distribution Wright-Patterson AFB	Repair Requirements Computation System, developed at Wright-Patterson AFB, operated at Kelly AFB
IBM 305 Base Supply Offut AFB	Base Level Inventory Control System Scott AFB
Engine Management System Tinker AFB	MILSTAMP Central Data Collection System McClellan AFB

A table of the ADP systems in the final sample is given in Table 2. The orientation of approach for management supporting systems (e.g., Base Supply System) and operations supporting systems (e.g., SPACETRACK) was toward "single application" as objects of interest. A single application is a set of programs dedicated to one function which operates on part or all of a hardware configuration. The research and development supporting systems in the sample are all "scientific job shops" where numerous single applications exist on the same machine. The approach at R&D installations was to select one of the many single applications as an object of interest.

F. Data Collection Procedures

1. Initial Design of Questionnaire

Initial design of the questionnaire to be used in field collection of data was based on the work of Phase I. Appendix I of the Phase I final report provided a partially designed questionnaire, which Phase II project members used as a point of departure upon which to apply modifications. Project members were assigned specific areas of the questionnaire according to their specialties. Thus, an individual with extensive experience in programming was assigned portions of the questionnaire relative to programming, while an individual with extensive experience in operations was assigned a questionnaire section dealing with computer operations. The questions were then brought together and organized to form a comprehensive questionnaire.

Some variables originally postulated in Phase I, such as overhead cost and facilities cost, were not included in the questionnaire. The reason these costs were left out was that the effort required to gather this type of cost data and place it on a uniform basis could be much more profitably spent on more central areas of the ADPS.

The initial questionnaire was based on the ADPS model described in subsection II.B and was directed toward the following individuals:

TABLE 2 - ADP SYSTEMS IN SAMPLE

Stratification	Small	Medium	Large
Management supporting data systems--logistics (similar Set I)	Base-level electronic inventory control system (GE/BSS), Scott AFB Data services workload control (DSWC), Kelly AFB	MILSTAMP Central Data Collection System (MILSTAMP), McClellan AFB Priority distribution system (PDS), Wright-Patterson AFB	Base supply system (1050/BSS), HQ USAF Repair requirement computation System (RRC), Wright-Patterson AFB
Management supporting data systems--personnel/finance (similar Set II)	Accrued military pay system (AMPS), accounting and finance center AFSC accounting (SC/ACCT), Andrews AFB	Merged accountability and fund reporting (MAFR), accounting and finance center Regional accounting and finance test (RAFT), Randolph AFB	PDS-0, Military Personnel Center (PDSO/MPC), Randolph AFB PDS-0, Major Air Command (PDSO/MAC), Randolph AFB
Operations supporting data systems (dissimilar Set I)	TAC Command and Control (TCC), Langley AFB	Global Weather Central (GWC), Offut AFB	Spacetrack (SPCTRK), Ent AFB
Research and development supporting systems (dissimilar Set II)	Project Adobe data reduction (ADOBE), Rocket Propulsion Laboratory, Edwards AFB	Orbit determination and analysis (ORBIT), Cambridge Research Laboratory, Hanscom Field	Missile simulation (MISSIM), Air Proving Ground Center, Eglin AFB

1. Installation Manager: To supply general organizational, functional, and historic information on system development, operation, and use.

2. Systems/Programming Supervisor: To supply information on development costs, documentation, support software, application software, workload descriptors, and personnel data for analysts/programmers.

3. Operations Supervisor: To supply information on job scheduling, computer utilization, hardware and facility problems, and personnel data for operators.

The initial questionnaire was highly detailed and of relatively fixed format; i.e., it was composed of numerous multiple choice and specifically directed questions and fixed tabular forms for recording such data as workload descriptors.

2. Modification of Questionnaire

To verify the usefulness of the initial questionnaire, a pilot data collection trip was made to Randolph AFB, Texas. The subject system was the Personnel Data System for Officers (PDSO-65), which operated on the Burroughs B5500 computer. This system was an excellent choice for pilot data collection, since it included a very broad spectrum of capabilities and features, such as large direct access memory, on-line inquiry capability, and multiprogramming. The breadth of this system ensured that a questionnaire which could handle it would be applicable to a wide variety of systems.

Although a wealth of data was available on PDSO-65, difficulty was encountered in placing this data in the rigid format of the initial questionnaire. During the Randolph pilot data collection, the questionnaire was modified to conform to the availability and type of data that was encountered, and the data were recorded on the modified questionnaire.

The availability of data at Randolph, particularly in the proposal and workload areas, suggested that highly reliable data would be available in all areas specified by Phase I. (Experience in subsequent data collection revealed that availability of data on PDSO-65 was very high.)

On returning from Randolph AFB, project personnel developed a more general questionnaire. The revised questionnaire assumed the format of an interviewer's guideline together with a number of tabular sheets for entering fixed information. A copy of the revised questionnaire is included in Appendix B.

3. Letter of Introduction

On 30 March 1966, a letter introducing the project to all installations to be interviewed was signed for Hewitt T. Wheless, Lt., General,

USAF, Assistant Vice Chief of Staff. This letter (see Figure 3) assisted significantly in the data collectors' receiving outstanding cooperation from the installations visited.

4. Data Collection Trips

In addition to the pilot data collection trip of 14 to 21 March 1966, three other series of trips have been made. The first series of trips covered the period 11 to 22 April, the second covered the period 9 to 20 May, and the last series was in July 1966. Each team had 2 weeks to cover two systems, except for one team during the second series, which spent 1 week on one system. This staffing was found to be adequate, and, in a number of cases, the requisite data were collected ahead of schedule. An average of about 8 man-days per system was required on site to collect data.

Between 1 and 2 weeks in advance of the data collection trips, the lead data collection team member contacted the AFADA-designated contact to inform him of the purpose of the trip and the type of data to be collected. Arrangements for time and place of meeting on arrival at the installation were also made. On arriving at an installation, project personnel briefed key installation personnel on the goals of the project and on the types of data to be collected. PRC, in turn, asked for a brief orientation defining organizational responsibility and general system characteristics.

The general order of data collection was as follows:

Day	Data Collection Activity
1st	Briefing, organizational, and functional description
2nd	System proposal, personnel, and manpower data
3rd	Programming and workload data
4th	Operations and workload data
5th	Review of all data and debriefing

The last task was always a debriefing for the key installation personnel to inform them of the data that had been collected. Appendix I provides a list of personnel contacted during each of the data collection trips.

G. Data Reduction

1. Data Summarization

On returning from a data collection trip, team members proceeded to reduce raw numeric data to consistent meaningful quantities,

DEPARTMENT OF THE AIR FORCE
OFFICE OF THE CHIEF OF STAFF
UNITED STATES AIR FORCE
WASHINGTON, D.C.

30 MAR 19[illegible]

REPLY TO ATTN OF: AFCCS

SUBJECT: Collection of Information on Automatic Data Processing Systems

TO: See Distribution

1. In 1964 the Secretary of the Air Force asked for a study of the best way to use Air Force Automatic Data Processing Systems (ADPS) experience in judging proposals for new automation. Two competitive contracts were awarded by the Electronic Systems Division of the Air Force Systems Command to develop approaches to solving this problem. As a result of the competition, Planning Research Corporation (PRC) was awarded a contract to collect data on 18 existing automatic data processing systems to test their proposed approach and to compile an experience compendium. The systems to be examined fall in the areas of Logistics, Personnel/Finance and Accounting, Command and Control and R&D Support. A list of them is attached.

2. This will be the first known attempt to develop general broad-spectrum techniques for estimating costs and development times for complete data processing systems at the proposal stage. If successful, the Air Force will receive significant benefits from the effort.

3. The data collection phase of the contract will extend through July 1966. Air Force and PRC management personnel will contact each facility on the attached list in advance of the data collection operation to brief on the project, answer questions, and arrange details of the later data collection visit.

4. In order to reduce the impact on your operations and to secure greater uniformity of data, all data collection will be done by PRC personnel. In general the data to be gathered will include:

a. A general description of the automatic data processing system and the organizations involved in its development and use.

b. The development schedule, both as planned and as realized.

c. The development and operating cost history.

FIGURE 3 - LETTER OF INTRODUCTION

d. Descriptions of each type of input record (with frequency of use), each type of output record (with response times), and each data-base file and record.

5. Request PRC be given access to records of systems on attached list and personnel associated with these systems.

6. Questions concerning this matter may be directed to AFRDQRC, ATTN: Lt Colonel John J. Hobson, extension 71029.

FOR

HEWITT T. WHELESS, Lt General, USAF
Assistant Vice Chief of Staff

1 Attachment
List of Systems to be Examined by PRC

FIGURE 3 (Continued)

which had been identified as the independent and dependent variables for statistical analysis. Examples of the numeric data reduction include computation of personnel experience averages, percentages of computer hours for different processing functions, and total characters/month of input volume.

A comprehensive system writeup was then prepared. This document, which ranged in length from 30 to 60 pages, presented a narrative and graphic description of organizational relationships, system history, proposal for the system, schedule, system design, documentation, programming, file conversion, operations, computer utilization, personnel, manpower, and future plans. The system writeup, together with the original file of raw data, serves as the basic source document for statistical analysis inputs, for the system summaries of the midpoint report, and for the system descriptions of the experience handbook.

The numeric and narrative data reduction required about 12 man-days per system, on the average.

The original plan for Phase II called for a file maintenance program to be developed for maintaining and reducing the raw data collected. This effort was declared unnecessary after the pilot data collection trip, when it appeared that raw data would not be uniform from system to system.

Summary sheets (see Appendix B) for the numerical data collected were prepared, and listed all dependent and independent variables. Variables for which data must be collected were indicated.

The data from the summary sheets of the 18 systems were compiled onto work sheets. Copies of the worksheets were provided each data collector for audit and recheck of the data he was responsible for collecting. Appendix D displays the worksheet and the raw data collected for each of the 18 systems. An indication of the reliability of the collected data is also given. An illustration of the total data summarization process is depicted in Figure 4.

2. <u>Quality Control of Data</u>

The previous discussion of operational definitions (subsection II.D) described the redefinition of variables following the second and third data collection trips. In addition to audit and recheck of data, another review session was held with the data collectors of each system. The purpose was to evaluate system data against the refined definitions. System data was assessed for accuracy and reliability, completeness, level of detail and precision, proper categorizations, and currency.

As a result of this review, some variables were dropped from the data analysis because of too many missing values. Examples are average frequency per hour of input and hours of compilation, assembly, checkout, and system test during the development phase. Because of

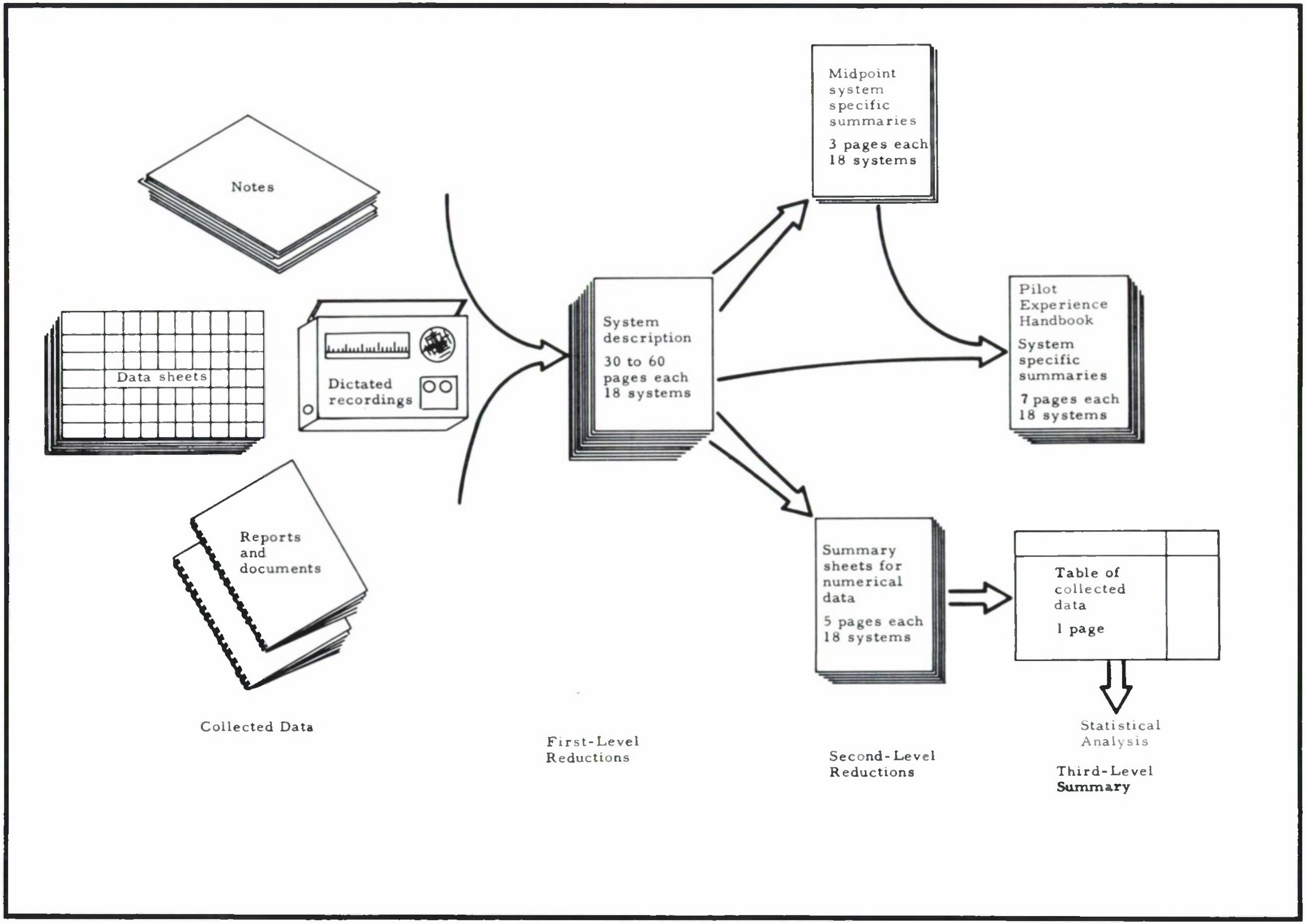


FIGURE 4 - DATA REDUCTION

the small sample size, the normal procedure of applying the mean of the variable for missing values may distort the results if too many (for example, more than two) missing values were substituted in this way. Furthermore, since there was a lack of "direct" input and output data (two systems had such data), the data for the variable were merged with batched inputs and outputs. "Direct" refers to on-line computer input and output without computer operator intervention. (See Appendix C for definitions of input and output variables $X_1, X_2, \ldots, X_6$. Other variables with data considered unreliable were also dropped from the analysis.

H. Findings

Workload and cost data for an ADPS were generally collectible and reducible, but reliability was not as high as if data were recorded at the time of event occurrence. The problems encountered were that workload and cost data often were not recorded or they were aggregated such that they had become inseparable.

Current ADPS proposals do not contain sufficient data about planned costs and contain nothing about workload descriptors. No proposals could be found for some of the older systems.

III. DATA ANALYSIS

This section discusses the objectives of, the procedures followed for, and the findings of the data analysis.

A. Objectives

The Phase II goals for data analysis were as follows:

- o To determine relationships between ADPS workload descriptors and cost and development time
- o To validate that the workload descriptors exhibit inclusiveness (intuitively similar systems have similar workloads), exclusiveness (intuitively dissimilar systems have significantly different workloads), and breadth of application (workload descriptors are applicable to a wide variety of systems)
- o To refine workload descriptors by defining them more precisely, by eliminating nonsignificant descriptors, and by combining significant descriptors

To achieve these objectives the following tasks were performed:

- o Refining the workload model using scatterplot analysis and correlation analysis
- o Testing the model using analysis of variance and analysis of covariance
- o Developing cost estimation equations using multiple regression analysis
- o Determining measures of reliability for the cost estimation equations
- o Using factor analysis to discover other potential relationships and to check the cost estimating relationships that were derived

Each of these tasks is described in the subsequent discussion. Finally, a section covering findings and conclusions completes this section on data analysis.

B. Workload Model Refinement

Following data reduction, the workoad model as described in Section II, Data Collection, was reduced from 87 to 43 variables. These

consisted of 26 independent variables, constituting the preliminary set of estimating factors, and 17 dependent variables, constituting the preliminary set of planning factors. These variables were carefully examined and classified into logical categories, as shown in Table 3 and defined in Appendix C. The independent variables are postulated as the causal factors and the dependent variables as the effects on the cost and other phenomena of ADP systems.

1. Selection of Factors To Be Estimated

The initial step in the refinement process was to determine the relevant planning factors in the model. These are the factors to be estimated. The dependent variables were critically scrutinized to select those factors that would be of value in the proposal judging process. The single criterion employed was that the variables chosen should consist of the minimum set that would include all development costs, all operations costs, and development time. The following factors were selected:

Cost Factors

Development cost variables

Y_1 Development effort

Y_2 Program checkout, hardware cost

Operations cost variables

Y_3 Program maintenance personnel

Y_4 Operations personnel

Y_5 Application production, hardware cost

Y_6 Program maintenance, hardware cost

Other Factors

Y_7 Elapsed development time

The factor $Y2$ (program checkout, hardware cost) had to be eliminated from the statistical analysis because of insufficient data; many ADP systems did not record this information separately during development. Source statements and object instructions, although interesting, are not included because it would be preferable to obtain development cost directly. Application production factors do not appear to be of significant import for proposal evaluation. As mentioned in subsection II.F.1, facilities costs were not included in the study. In practice, these costs may be estimated by the computer size and the number of personnel.

2. Selection of Factors To Be Used as Predictors

The next step in the refinement process was a determination of relevant predictors available in the model. These predictors will be used in estimating the relevant planning factors. The independent

TABLE 3 - CLASSIFICATION OF VARIABLES

Dependent Variables (Planning Factors)

Class	Symbol	Name	Units of Measurement
Cost Factor			
Development cost variables	Y_1	Development effort	Man-months
	Y_2	Program checkout, hardware cost[1]	Dollars
Operations cost variables	Y_3	Program maintenance personnel	Number
	Y_4	Operations personnel	Number
	Y_5	Application production, hardware cost[1]	Dollars/month
	Y_6	Program maintenance, hardware cost[1]	Dollars/month
Other Factors	Y_7	Elapsed development time	Months
	Y_8	Source statements	Number
	Y_9	Object instructions	Number
Application Production Factor	Y_{10}	Input edit, production hours	Percent
	Y_{11}	File maintenance, production hours	Percent
	Y_{12}	Report generation, production hours	Percent
	Y_{13}	Merge, production hours	Percent
	Y_{14}	Sort, production hours	Percent
	Y_{15}	Compute, production hours	Percent
	Y_{16}	Query, production hours	Percent
	Y_{17}	Control, production hours	Percent

Independent Variables (Estimating Factors)

Class	Symbol	Name	Units of Measurement
Workload Factor			
Input variable	X_1	Input volume	Characters/month
	X_2	Input transaction types	Number
	X_3	Input data fields	Number
	X_4	Input rejects	Percent
Output variable	X_5	Output volume	Characters/month
	X_6	Output formats	Number
Data base variable	X_7	Data base	Characters
	X_8	Data base record types	Number
Complexity Factor	X_9	Input edit, source statements	Percent
	X_{10}	File maintenance, source statements	Percent
	X_{11}	Report generation, source statements	Percent
	X_{12}	Merge, source statements	Percent
	X_{13}	Sort, source statements	Percent
	X_{14}	Compute, source statements	Percent
	X_{15}	Query, source statements	Percent
	X_{16}	Control, source statements	Percent
Education and Experience Factor			
College education variable	X_{17}	Development managers	Average number of years
Automatic data processing experience variable	X_{18}	Development managers	Average number of years
	X_{19}	Analysts	Average number of years
	X_{20}	Programmers	Average number of years
	X_{21}	Operations personnel	Average number of years
Functional area experience variable	X_{22}	Development managers	Average number of years
	X_{23}	Analysts	Average number of years
	X_{24}	Programmers	Average number of years
	X_{25}	Operations personnel	Average number of years
Machine Maturity Factor	X_{26}	Machine maturity	Months

Note: (1) See Appendix E for method of computing hardware cost.

variables were critically scrutinized to select those factors that are causally related to the planning factors and that can be available at proposal evaluation time from the ADPS proposal. The latter criterion was considered to be of paramount importance because estimating relationships that were derived from nonavailable factors cannot be put to use. The following factors were selected:

Workload Descriptors

Input variables

X_1 Input volume

X_2 Input transaction types

X_3 Input data fields

Output variables

X_5 Output volume

X_6 Output formats

Data base variables

X_7 Data base (size)

X_8 Data base record types

The factor X_4 (input rejects) was not included because it may not be determinable at proposal preparation time. While the complexity factor, education and experience factor, and machine maturity factor are important, they are not usually known at the proposal preparation time. Complexity is difficult to determine until system design is nearly complete; the machine (computer) to be used is usually acquired subsequent to proposal approval; and education and experience levels of personnel are largely unknown until the proposal is being implemented and staffing largely completed.

3. General Linear Model

The foundation for the use of regression analysis is establisted in this subsection and in the one immediately following. These subsections do not pertain specifically to the subject at hand and may be omitted by readers who are not interested in statistical methods.

It is assumed that a linear stochastic relationship exists between the dependent variables Y_j (cost variables) and a set of independent variables $X_1, X_2, \ldots, X_m$ (workload descriptors) and that this relationship may be written as

$$Y_j = \alpha_j + \sum_{i=1}^{m} \beta_{ij} X_i + U_j ,$$

where $j = 1, 2, \cdots, n$ defines the set of cost variables, and the frequency distribution of error terms α_j, $\beta 1_j$, $\beta 2_j$, $\cdots$, βm_j, and $f(U_j)$, define the data generating mechanism. The problem is to specify the dependency relationship correctly by selecting a proper set of independent variables, a proper functional form, and a vector of parameters

a_j, b_{1j}, b_{2j}, . . . , b_{mj} that provides a good set of estimates for α_j, β_{1j}, β_{2j}, . . . , β_{mj}, the structural coefficients of the underlying population (see Reference 10).

The estimates a_1, b_{1j}, b_{2j}, . . . , b_{mj} are obtained by regression analysis, and U_j's are determined and evaluated by measures of reliability. Both topics will be covered subsequently in this section.

4. Requirements for Estimation Efficiency

Estimation efficiency means the accuracy of the estimating relationship. The estimation efficiency of the prediction equation derived by means of regression analysis is conditioned on the following requirements:

1. $E(Y_j) = a_j + \Sigma b_{ij}X_i$ is linear in the specified set of parameters and independent variables.

2. $f(U_j)$ is normal; the conditional distribution of Y_1 given X_1, X_2, . . . , X_m follows the normal probability function.

3. $E(U_jU_{j+s}) = 0$ for all $S \neq a$; successive errors are independently distributed.

4. $E(U_j^2)$ is constant (the variance of error terms is independent of the size of explanatory variables X_i); absence of heteroscedasticity (heteroscedasticity refers to the nonuniformity of the variance of the Y variable through the range of the X variable).

5. $E(X_i X_k) = 0$; independent variables are independent of one another (absence of multicollinearity).

6. $E(U_jX_i) = 0$, for all $i = 1, 2, \ldots, m$; the requirement that Y be dependent on X but not vice-versa (absence of feedback).

These requirements were checked by the use of scatterplot analysis and correlation analysis, which will be covered in subsequent subsections. The residuals appeared normal when plotted as transformed.

5. Scatterplot Analysis

The purpose of the scatterplot was to provide for visual inspection of the distribution of variables. (See subsection B of Appendix E for the methodology used for generating scatterplots.) An examination of the plots of independent variables against the dependent variables showed tight clustering of data points with an extreme outlier or a fanning out of data at the higher ends of each scale (see Figure 5). This indicated a lack of linearity of the Y and X relationship and nonuniformity

of the Y variable through the range of the X variable, which violate requirements for the use of regression analysis.

In attempting to improve the distribution of data, transformation of variables was performed. Logarithmic (base 10) transformation of the independent variable only, $\log_{10}$ transformation of the dependent variable only, and $\log_{10}$ transformation of both variables were tried. It was found that $\log_{10}$ transformation of both variables was necessary to produce a resulting distribution that appeared rectilinear and suitable for analysis. Figure 6 shows the results of $\log_{10}$ transformation on the same variables shown in Figure 5. (See subsection C of Appendix E for the methodology used for transformation of variables.)

Transformation techniques are intended to manipulate the data so that the resulting distribution of data will match the assumptions demanded for using linear regression analysis. The use of transformation is not intended to improve the results nor the reliability of the estimation equations derived. Furthermore, while all computations to determine reliability are performed in the transformed state, the ultimate estimates must be retransformed before use.

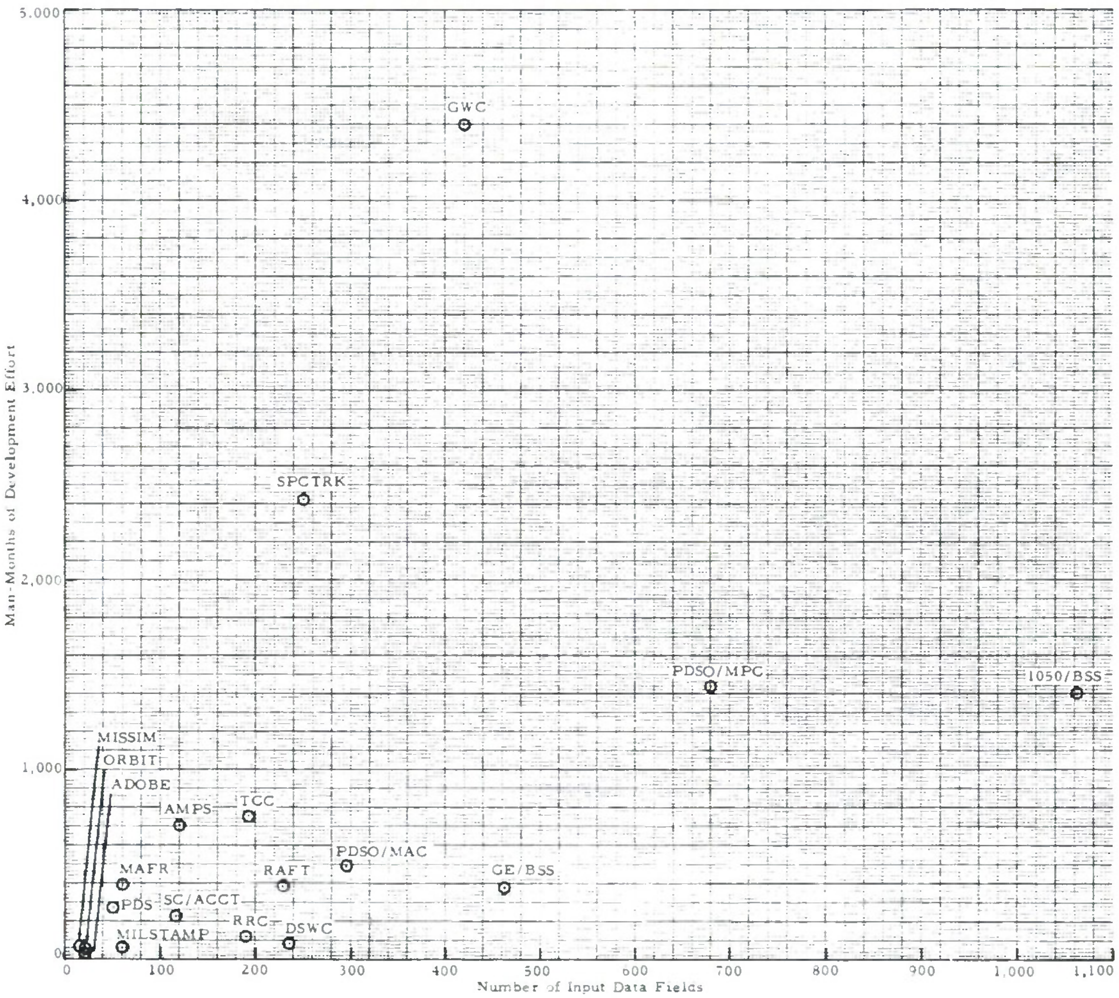
Figures 6 through 10 display examples of scatterplots. These show the good correlations between the primary workload descriptor against each of the five cost variables. (See also Table 4.)

6. Correlation Analysis

A correlation matrix for all transformed independent and dependent variables was computed. (See subsection D in Appendix E for methodology used for correlation.) This matrix was closely examined, and each cost variable Y_j is postulated to be a function of all workload descriptors X_i that are significantly correlated with Y_j for that given sample size.

The causal relationships between each Y_j and remaining X_i's were analyzed to ensure that the effect of Y_j is caused by the X_i's but that Y_j has no cause-effect relationships with the X_i's. None of the latter relationship was found.

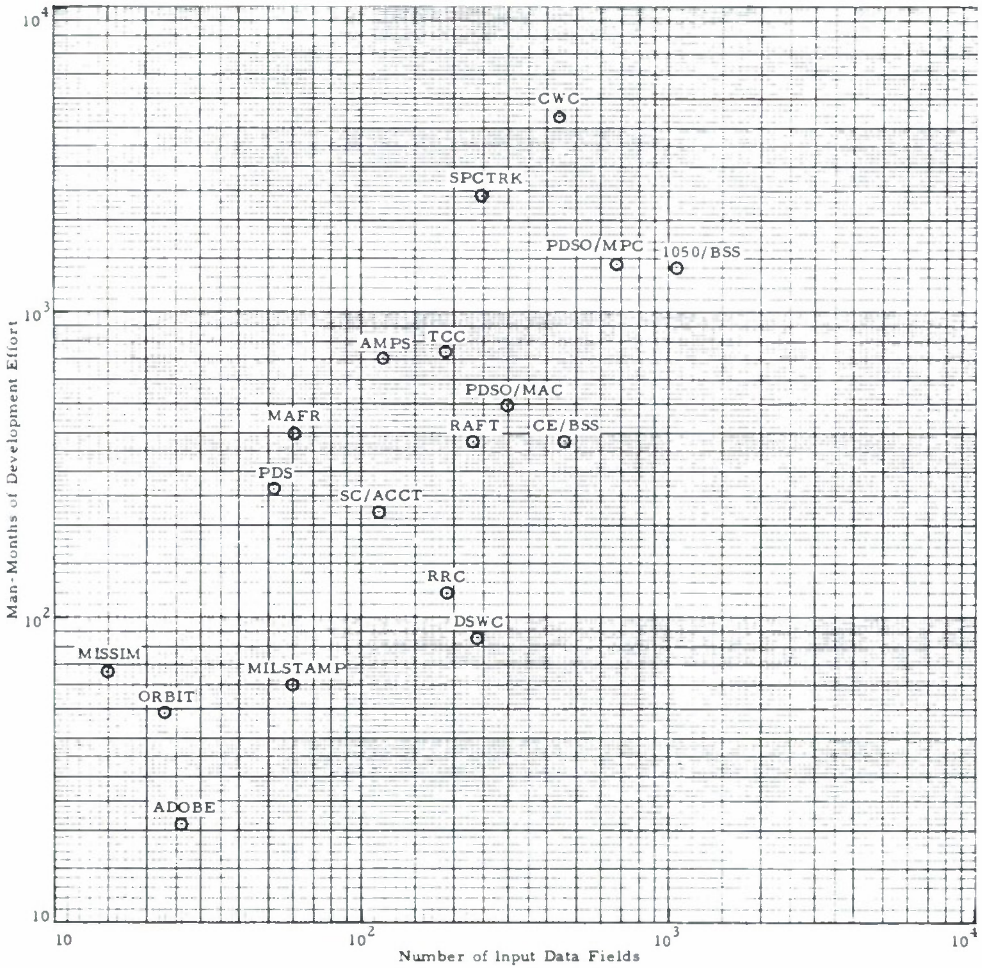
Finally, the X_i's for each Y_j were examined for intercorrelations. When significant intercorrelations existed, one or more variables were deleted until there was an absence of intercorrelated variables. The results of the correlation analysis are summarized in the cost model shown in Table 4. The regression model (the model entering the regression analysis) contained only the variables remaining following the elimination process previously described. The planning factor, elapsed development time (Y_7), was dropped because of lack of significant correlation with any workload descriptor. This was probably because of small sample size that did not supply sufficient data points to establish this relationship. Elapsed development time may still be estimated as a function of the number of man-months of development effort.



This plot displays the data point clustering and extreme value effect resulting in a poor correlation, 0.44, between man-months of development effort, and the number of input data fields when the data are not transformed.

Acronym	ADP System
ADOBE	Project Adobe Data Reduction
AMPS	Accrued Military Pay System
DSWC	Data Services Workload Control
GE/BSS	Base-Level Inventory System
GWC	Global Weather Control
MAFR	Merged Accountability and Fund Reporting III, MAFR III
MILSTAMP	Air Force MILSTAMP Central Data Collection
MISSIM	Missile Simulation
ORBIT	Orbit Determination and Analysis
PDS	Priority Distribution System
PDSO/MAC	Major Air Command Personnel Data System--Officers 65
PDSO/MPC	Personnel Data System--Officers, Military Personnel Center
RAFT	Regional Accounting and Finance Test
RRC	Repair Requirement Computation System
SC/ACCT	Appropriation Accounting Remote Random Access System
SPCTRK	SPACETRACK
TCC	TAC Command and Control
1050/BSS	Standard Base-Level Automated Inventory Control System

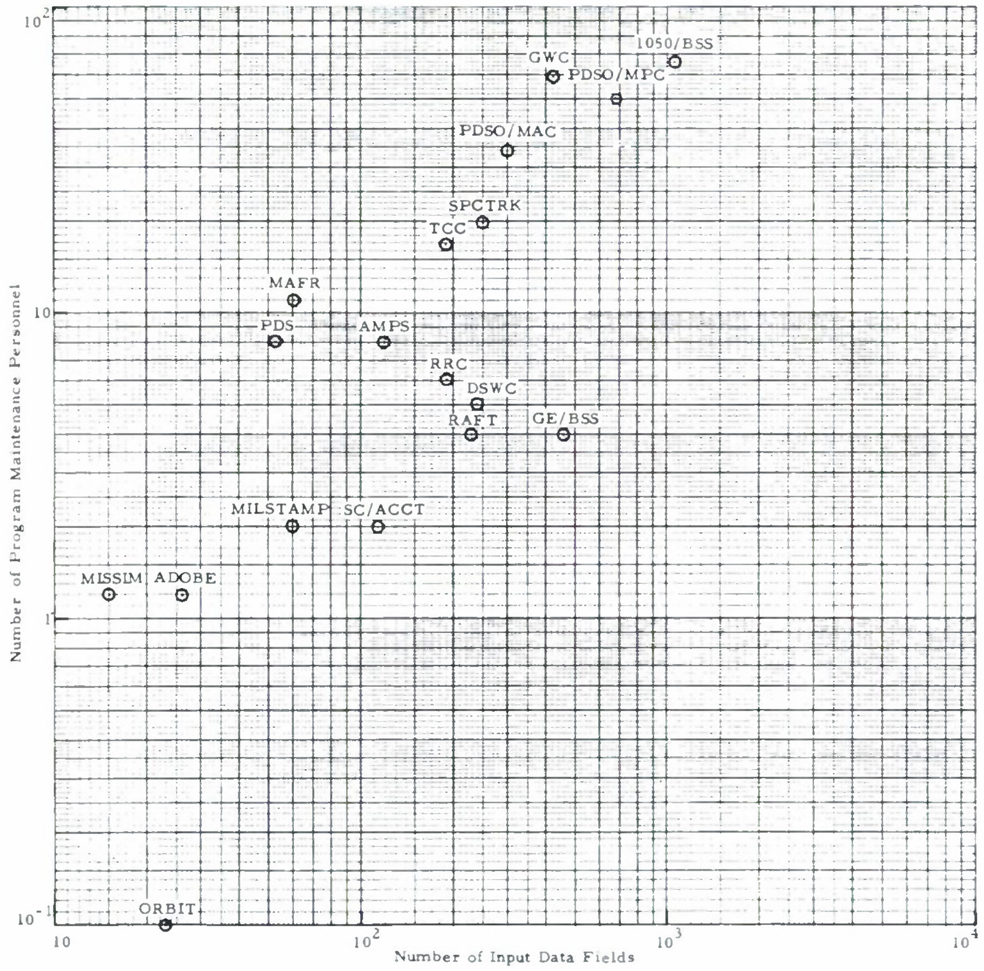
FIGURE 5 - MAN-MONTHS OF DEVELOPMENT EFFORT VERSUS NUMBER OF INPUT DATA FIELDS (BEFORE LOG_{10} TRANSFORMATION)



This plot displays the good correlation, 0.75, between man-months of development effort and the number of input data fields transformed to $\log_{10}$.

Acronym	ADP System
ADOBE	Project Adobe Data Reduction
AMPS	Accrued Military Pay System
DSWC	Data Services Workload Control
GE/BSS	Base-Level Inventory System
GWC	Global Weather Control
MAFR	Merged Accountability and Fund Reporting III, MAFR III
MILSTAMP	Air Force MILSTAMP Central Data Collection
MISSIM	Missile Simulation
ORBIT	Orbit Determination and Analysis
PDS	Priority Distribution System
PDSO/MAC	Major Air Command Personnel Data System--Officers 65
PDSO/MPC	Personnel Data System--Officers, Military Personnel Center
RAFT	Regional Accounting and Finance Test
RRC	Repair Requirement Computation System
SC/ACCT	Appropriation Accounting Remote Random Access System
SPCTRK	SPACETRACK
TCC	TAC Command and Control
1050/BSS	Standard Base-Level Automated Inventory Control System

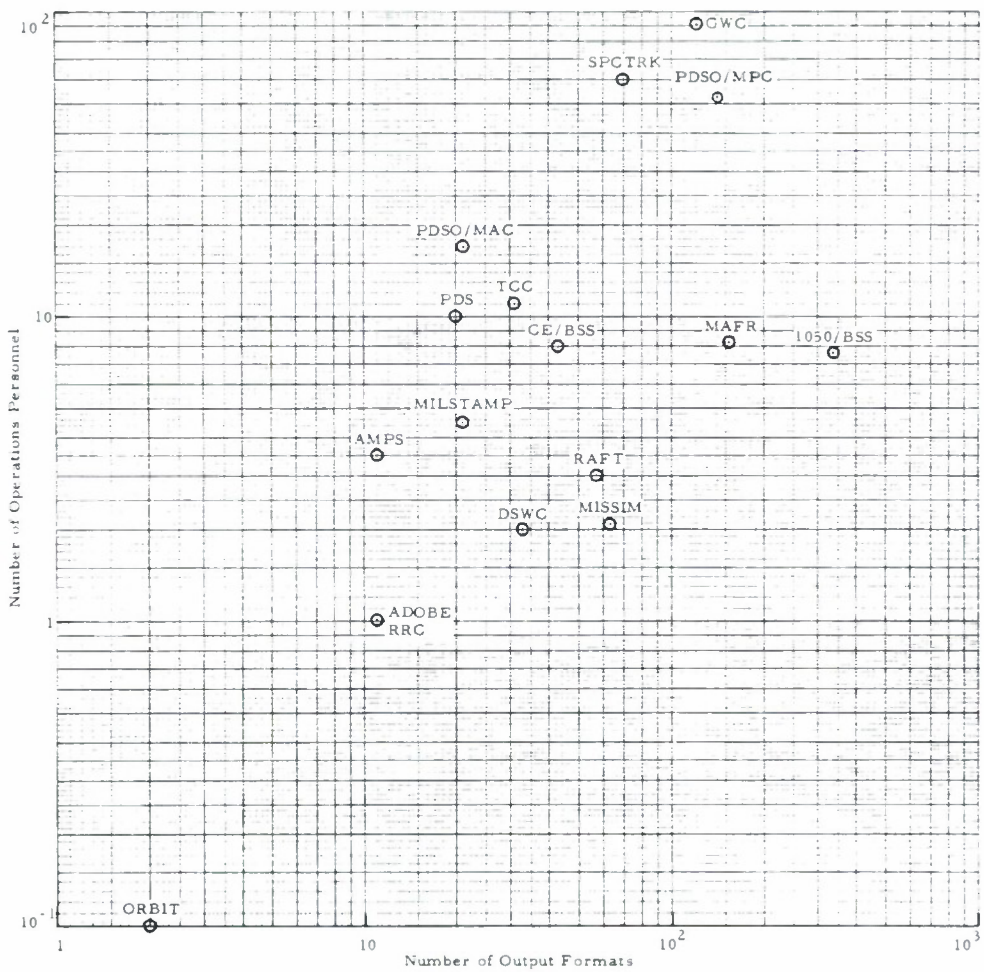
FIGURE 6 - MAN-MONTHS OF DEVELOPMENT EFFORT VERSUS NUMBER OF INPUT DATA FIELDS



This plot displays the good correlation, 0.78, between the number of program maintenance personnel and the number of input data fields transformed to $\log_{10}$.

Acronym	ADP System
ADOBE	Project Adobe Data Reduction
AMPS	Accrued Military Pay System
DSWC	Data Services Workload Control
GE/BSS	Base-Level Inventory System
GWC	Global Weather Control
MAFR	Merged Accountability and Fund Reporting III, MAFR III
MILSTAMP	Air Force MILSTAMP Central Data Collection
MISSIM	Missile Simulation
ORBIT	Orbit Determination and Analysis
PDS	Priority Distribution System
PDSO/MAC	Major Air Command Personnel Data System--Officers 65
PDSO/MPC	Personnel Data System--Officers, Military Personnel Center
RAFT	Regional Accounting and Finance Test
RRC	Repair Requirement Computation System
SC/ACCT	Appropriation Accounting Remote Random Access System
SPCTRK	SPACETRACK
TCC	TAC Command and Control
1050/BSS	Standard Base-Level Automated Inventory Control System

FIGURE 7 - NUMBER OF PROGRAM MAINTENANCE PERSONNEL VERSUS NUMBER OF INPUT DATA FIELDS



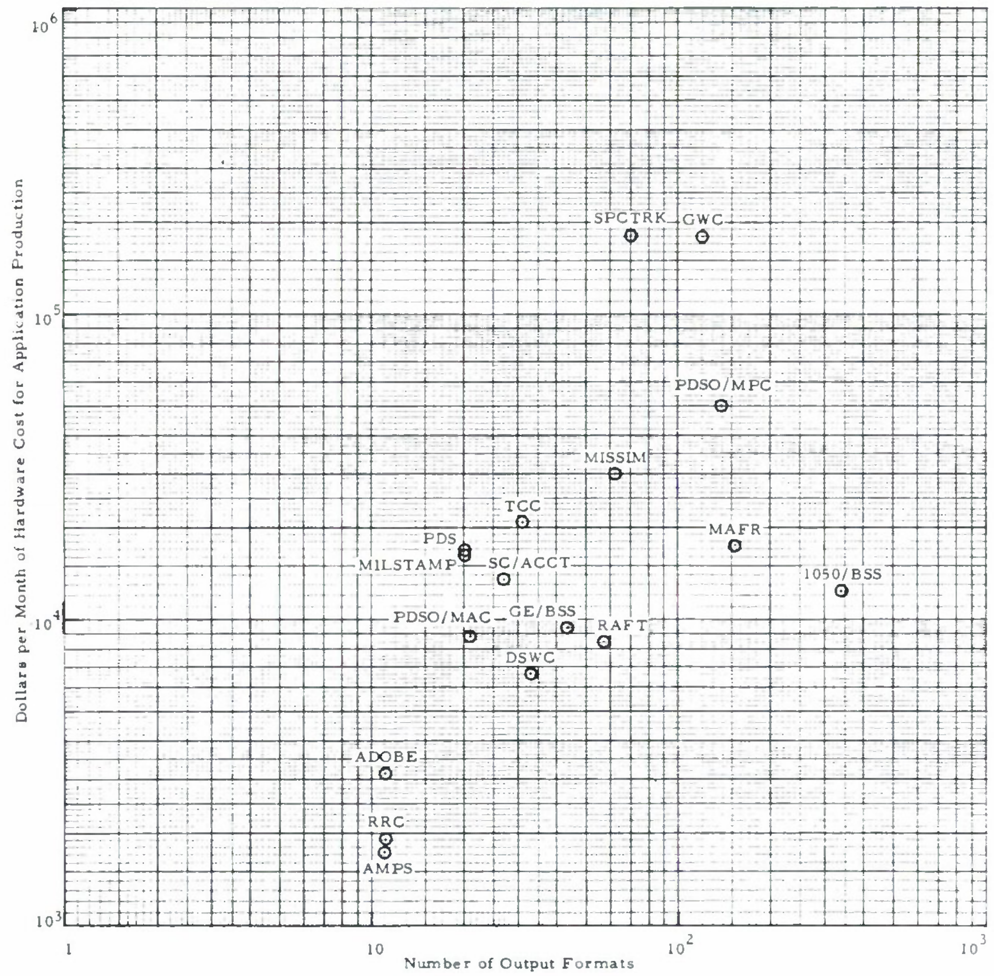
This plot displays the good correlation, 0.71, between the number of operations personnel and the number of output formats transformed to $\log_{10}$.

Acronym	ADP System
ADOBE	Project Adobe Data Reduction
AMPS	Accrued Military Pay System
DSWC	Data Services Workload Control
GE/BSS	Base-Level Inventory System
GWC	Global Weather Control
MAFR	Merged Accountability and Fund Reporting III, MAFR III
MILSTAMP	Air Force MILSTAMP Central Data Collection
MISSIM	Missile Simulation
ORBIT	Orbit Determination and Analysis
PDS	Priority Distribution System
PDSO/MAC	Major Air Command Personnel Data System--Officers 65
PDSO/MPC	Personnel Data System--Officers, Military Personnel Center
RAFT	Regional Accounting and Finance Test
RRC	Repair Requirement Computation System
SC/ACCT	Appropriation Accounting Remote Random Access System
SPCTRK	SPACETRACK
TCC	TAC Command and Control
1050/BSS	Standard Base-Level Automated Inventory Control System

Points Not Plottable

ADPS	Number of Output Formats	Number of Operations Personnel
SC/ACCT	27	Unknown

FIGURE 8 - NUMBER OF OPERATIONS PERSONNEL VERSUS NUMBER OF OUTPUT FORMATS



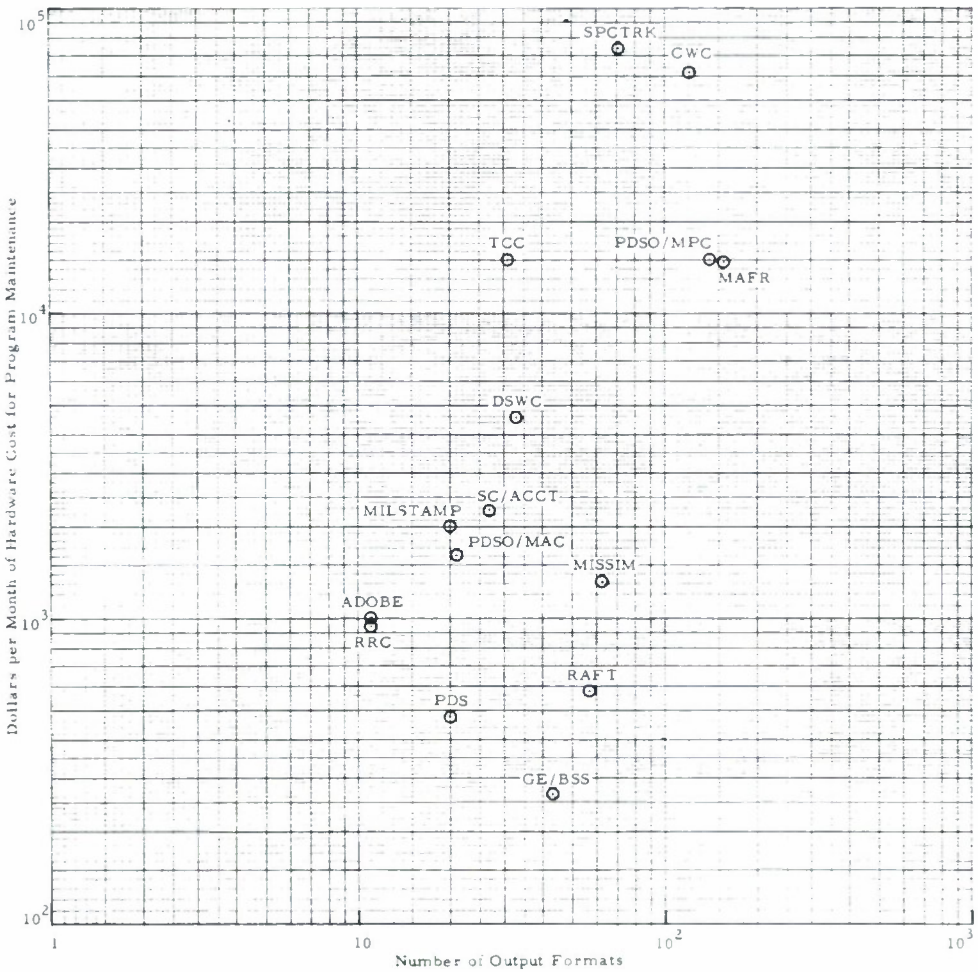
This plot displays the good correlation, 0.77, between dollars per month of hardware cost for application production and the number of output formats transformed to $\log_{10}$.

Acronym	ADP System
ADOBE	Project Adobe Data Reduction
AMPS	Accrued Military Pay System
DSWC	Data Services Workload Control
GE/BSS	Base-Level Inventory System
GWC	Global Weather Control
MAFR	Merged Accountability and Fund Reporting III, MAFR III
MILSTAMP	Air Force MILSTAMP Central Data Collection
MISSIM	Missile Simulation
ORBIT	Orbit Determination and Analysis
PDS	Priority Distribution System
PDSO/MAC	Major Air Command Personnel Data System--Officers 65
PDSO/MPC	Personnel Data System--Officers, Military Personnel Center
RAFT	Regional Accounting and Finance Test
RRC	Repair Requirement Computation System
SC/ACCT	Appropriation Accounting Remote Random Access System
SPCTRK	SPACETRACK
TCC	TAC Command and Control
1050/BSS	Standard Base-Level Automated Inventory Control System

Points Not Plottable

ADPS	Number of Output Formats	Dollars/Month of Cost for Application Production
ORBIT	2	140

FIGURE 9 - DOLLARS PER MONTH OF HARDWARE COST FOR APPLICATION PRODUCTION VERSUS NUMBER OF OUTPUT FORMATS



This plot displays the good correlation, 0.76, between dollars per month of hardware cost for program maintenance and the number of output formats transformed to $\log_{10}$.

Acronym	ADP System
ADOBE	Project Adobe Data Reduction
AMPS	Accrued Military Pay System
DSWC	Data Services Workload Control
GE/BSS	Base-Level Inventory System
GWC	Global Weather Control
MAFR	Merged Accountability and Fund Reporting III, MAFR III
MILSTAMP	Air Force MILSTAMP Central Data Collection
MISSIM	Missile Simulation
ORBIT	Orbit Determination and Analysis
PDS	Priority Distribution System
PDSO/MAC	Major Air Command Personnel Data System--Officers 65
PDSO/MPC	Personnel Data System--Officers, Military Personnel Center
RAFT	Regional Accounting and Finance Test
RRC	Repair Requirement Computation System
SC/ACCT	Appropriation Accounting Remote Random Access System
SPCTRK	SPACETRACK
TCC	TAC Command and Control
1050/BSS	Standard Base-Level Automated Inventory Control System

Points Not Plottable

ADPS	Number of Output Formats	Dollars/Month of Cost for Program Maintenance
1050/BSS	339	Unknown
AMPS	11	0
ORBIT	2	0

FIGURE 10 - DOLLARS PER MONTH OF HARDWARE COST FOR PROGRAM MAINTENANCE VERSUS NUMBER OF OUTPUT FORMATS

TABLE 4 - COST MODEL

Workload Descriptors (independent variables) / Cost Factors (dependent variables)		Log_{10} X_1 Characters Per Month of Input Volume	Log_{10} X_2 Number of Input Transaction Types	Log_{10} X_3 Number of Input Data Fields	Log_{10} X_5 Characters Per Month of Output Volume	Log_{10} X_6 Number of Output Formats	Log_{10} X_7 Characters in Data Base	Log_{10} X_8 Number of Data Base Record Types	Sample Size	H: $\rho = 0$(1) 0.05	0.01
Log_{10} Y_1	Man-months of development effort		0.61 X	0.75 ⊠		0.64 ⊠		0.72 X	18	0.456	0.575
Log_{10} Y_3	Number of program maintenance personnel		0.64 X	0.78 ⊠		0.73 ⊠		0.76 X	18	0.456	0.575
Log_{10} Y_4	Number of operations personnel	0.60 X			0.68 ⊠	0.71 ⊠	0.68 ⊠	0.64 X	17	0.468	0.590
Log_{10} Y_5	Dollars per month of hardware cost for application production	0.66 ⊠			0.67 X	0.77 ⊠	0.46 X	0.45 X	18	0.456	0.575
Log_{10} Y_6	Dollars per month of hardware cost for program maintenance	0.62 ⊠	0.40 X	0.39 X	0.60 X	0.76 ⊠		0.44 X	17	0.468	0.590

Note: (1) In order to say with 95 or 99 percent confidence that the true correlation coefficient is greater than zero, the sample correlation coefficient must be greater than the values given under the hypothesis $\rho = 0$.

Legend

X indicates variables with causal relationship

⊠ indicates variables with causal relationship which appear in cost estimation equations

0.79 correlation coefficient

Correlation Coefficient Ranking

> .9	Excellent
.8 - .9	Very Good
.7 - .8	Good
.6 - .7	Fair
< .6	Poor

C. Testing the Model

The ADPS sample was stratified into four subsamples (see Table 2). These were (1) management supporting data systems--logistics; (2) management supporting data systems--personnel/finance, (3) operations supporting data systems, and (4) research and development supporting systems. The subsamples were tested to determine whether their cost variables belong in the same population or should be treated as independent. Three tests were employed:

1. t-test to determine whether each of two subsamples belong in the same population.

2. One-way analysis of variance to determine whether all four subsamples belong in the same population.

3. Analysis of covariance to determine, after adjusting for differences in the workload descriptors, whether all four subsamples belong in the same population.

These tests were conducted for all five cost variables; the results are summarized in Table 5. (See subsection E in Appendix E for methodology used for analysis of variance.) The results were mixed and the 18 systems were treated as members of the same population.

Because relationships exist between workload descriptors and cost variables as demonstrated by the regression model, it should follow that, if the preceding tests were all significant, then the hypothesis that workload descriptors exhibit inclusiveness and exclusivness would be validated. The general applicability of workload descriptors has proved that they exhibit breadth of application. However, the results of the tests were mixed, and the hypothesis was not validated. One-way analysis of variance was also performed on the workload descriptors. The results were also mixed; however, these results do not mean that workload descriptors do not exhibit inclusiveness and exclusiveness. The mixed results are very likely due to the extremely small subsample sizes.

D. Regression Analysis

Multiple regression analysis using the stepwise regression procedure was applied to the cost models and five cost estimation equations were derived. Reliability measures (coefficients of variations and coefficients of correlation) and equations for the 80 percent prediction intervals were computed for each of the five estimating relationships. The equations are summarized in Table 6. (The statistical procedure is described in subsection F of Appendix E.)

For each cost variable, three cost estimates are obtained by using the appropriate workload descriptors and then by solving the prediction (cost estimation) equation and the prediction interval equation. The

TABLE 5 - RESULTS OF t-TESTS, ONE-WAY ANALYSIS OF VARIANCE, AND ANALYSIS OF COVARIANCE

Cost Factor	t-Test (Hypothesis: $\mu_i = \mu_j$)				One-Way Analysis of Variance (Hypothesis: $\mu_1 = \mu_2 = \mu_3 = \mu_4$)	Analysis of Covariance (Hypothesis: Means are equal after adjusting for differences in independent variables)
		μ_2	μ_3	μ_4		
Man-Months of Development Effort	μ_1	NS	0.05	0.05	0.05	0.01
	μ_2		0.05	0.05		
	μ_3			0.05		
Number of Program Maintenance Personnel	μ_1	NS	0.05	NS	0.05	NS
	μ_2		0.05	NS		
	μ_3			0.05		
Number of Operations Personnel	μ_1	NS	0.05	0.05	0.05	NS
	μ_2		0.05	NS		
	μ_3			0.05		
Dollars per Month of Hardware Cost for Application Production	μ_1	NS	NS	0.05	0.05 significant 0.01 NS	NS
	μ_2		NS	0.05		
	μ_3			NS		
Dollars per Month of Hardware Cost for Program Maintenance	μ_1	NS	NS	0.05	NS	NS
	μ_2		NS	0.05		
	μ_3			NS		

Notes: (1) NS--Not significant; means accept hypothesis

(2) 0.05--Means reject hypothesis at $\alpha = 0.05$, 95 percent level of confidence

(3) 0.01--Means reject hypothesis at $\alpha = 0.01$, 99 percent level of confidence

(4) Subscripts on μ: 1--Management supporting data systems--logistics

2--Management supporting data systems--personnel/finance

3--Research and development supporting systems

4--Operations supporting data systems

TABLE 6 - COST ESTIMATION EQUATIONS

Equation Number	Prediction Equation	Multiple Correlation	Standard Error of Estimate	Coefficient of Variation	80 Percent Prediction Interval for Log $\hat{Y}$
I	$\hat{Y}_1 = 0.426 + 0.692\ X_3 + 0.376\ X_6$	0.79	0.41	0.16	$\hat{Y} \pm 0.554 \left[1.055 + 0.321\ (X_3 - 2.149)^2 + 0.311\ (X_6 - 1.544)^2 - 0.360\ (X_3 - 2.149)\ (X_6 - 1.544)\right]^{1/2}$
II	$\hat{Y}_3 = -1.68 + 0.757\ X_3 + 0.562\ X_6$	0.86	0.40	0.48	$\hat{Y}_3 \pm 0.530 \left[1.055 + 0.321\ (X_3 - 2.149)^2 + 0.311\ (X_6 - 1.544)^2 - 0.360\ (X_3 - 2.149)\ (X_6 - 1.544)\right]^{1/2}$
III	$\hat{Y}_4 = -2.59 + 0.292\ X_5 + 0.414\ X_6 + 0.0923\ X_7$	0.86	0.41	0.55	$\hat{Y}_4 \pm 0.560 \left[1.059 + 0.127\ (X_5 - 7.369)^2 + 0.375\ (X_6 - 1.551)^2 + 0.223\ (X_7 - 5.829)^2 - 0.234\ (X_5 - 7.369)\ (X_6 - 1.551) - 0.00556\ (X_5 - 7.369)\ (X_7 - 5.829) - 0.0342\ (X_6 - 1.551)\ (X_7 - 5.829)\right]^{1/2}$
IV	$\hat{Y}_5 = 0.732 + 0.297\ X_1 + 0.801\ X_6$	0.83	0.43	0.11	$\hat{Y}_5 \pm 0.578 \left[1.055 + 0.104\ (X_1 - 6.979)^2 + 0.288\ (X_6 - 1.544)^2 - 0.180\ (X_1 - 6.979)\ (X_6 - 1.544)\right]^{1/2}$
V	$\hat{Y}_6 = -3.353 + 0.458\ X_1 + 2.11\ X_6$	0.79	1.09	0.36	$\hat{Y}_6 \pm 1.47 \left[1.059 + 0.119\ (X_1 - 6.985)^2 + 0.423\ (X_6 - 1.486)^2 - 0.270\ (X_1 - 6.985)\ (X_6 - 1.486)\right]^{1/2}$

$\hat{Y}_1 = \log_{10}$ (predicted man-months of development effort)
$\hat{Y}_3 = \log_{10}$ (predicted number of program maintenance personnel)
$\hat{Y}_4 = \log_{10}$ (predicted number of operations personnel)
$\hat{Y}_5 = \log_{10}$ (predicted dollars per month of hardware cost for application production)
$\hat{Y}_6 = \log_{10}$ (predicted dollars per month of hardware cost for program maintenance)

$X_1 = \log_{10}$ (characters per month of input volume)
$X_3 = \log_{10}$ (number of input data fields)
$X_5 = \log_{10}$ (characters per month of output volume)
$X_6 = \log_{10}$ (number of output formats)
$X_7 = \log_{10}$ (characters in data base)

prediction equation gives a single-point cost estimate that may be interpreted in the following manner: The cost is expected to be this value; 50 percent of the time it is expected to be greater, and 50 percent of the time it is expected to be less. The prediction interval provides a range estimate of cost. The solution of the prediction interval equation gives upper and lower limits that bracket the expected value and that may be interpreted in the following manner: Cost is expected to be less than the upper limit 90 percent of the time and greater than the lower limit 90 percent of the time.

For the Experience Handbook, these equations will be pre-solved with graphical methods. Figure 13 displays three iso-graphs representing Equation I.

Multiple regression analysis was also applied to the 12 management-supporting data systems. The resulting estimation equations for the five cost factors possessed prediction intervals of greater magnitude than the equations developed for all 18 systems. Again, the smaller sample size was a major contributor to the larger variance.

The estimating equations displayed wide prediction intervals. As an example, using Equation I with workload descriptors of 100 for the number of input data fields and 10 for the number of output formats, the solution obtained is 153 for the number of estimated man-months of development effort with a prediction interval of 38 to 614 man-months. With an increase in sample size from 18 to N, the interval width would decrease by a factor of $\sqrt{18/N}$ of the $\log_{10}$ values of the interval; if N = 180 , the interval limits will be 99 to 238 man-months.

The regression analyses undertaken all assumed a linear model with $\log_{10}$ transformed variables. Because of the small sample size, it was judged that polynominal regression would not provide estimating equations of better efficiency; therefore, it was not used.

E. Factor Analysis

Investigations leading toward development of a cost model incorporating complexity factors and personnel factors were not fruitful. Factor analysis was used in an attempt to discover potential significant relationships between these variables. (See subsection G in Appendix E for the methodology used in factor analysis.)

Of the 43 variables in the original reduced workload model, 31 were entered into the factor analysis; 12 were not entered because of insufficient data points. Only those variables with 17 or 18 observations were retained because of the requirements for uniformity in sample size. The eliminated variables had between 11 and 14 observations each, with one having 16 observations. The following variables were not entered:

Dependent Variables

Y_2 Hardware cost for program checkout

Y_{10} Percent of production hours for input edit

Y_{11} Percent of production hours for file maintenance

Y_{12} Percent of production hours for report generation

Y_{13} Percent of production hours for merge

Y_{14} Percent of production hours for sort

Y_{15} Percent of production hours for compute

Y_{16} Percent of production hours for query

Y_{17} Percent of production hours for control

Independent Variables

X_{17} Years of college education for development managers

X_{18} Years of ADP experience for operations personnel

X_{25} Years of functional area experience for operations personnel

Eight of the omitted dependent variables (Y_{10} through Y_{17}) were not significant with respect to ultimate utility because they were not cost variables. Therefore, their loss did not detract substantially from the analysis.

The intermediate results showed 17 nonzero eigenvalues and subsequently the factor matrix was rotated 17 times. Interpretation of the results of this sample following rotation of the factor matrix did not indicate that significant relationships existed between complexity variables and cost variables, and between education and experience variables and cost variables. The factor analysis did indicate, however, that significant relationships existed between the workload descriptors and cost factors of the cost estimating equations. This result was consistent with the results obtained previously. Factor loadings for selected variables from two of the 17 rotations of the factor matrix are displayed in Table 7. The high factor loadings are underlined. None of the other rotations showed any significantly high factor loadings in one or more dependent variables with one or more independent variables.

TABLE 7 - FACTOR LOADINGS FOR SELECTED VARIABLES

Independent Variables			
Symbol	Name	Factor I	Factor II
X_1	Characters per month of input volume	0.73	0.03
X_2	Number of input transaction types	0.18	0.68
X_3	Number of input data fields	0.21	0.70
X_5	Characters per month of output volume	0.72	0.02
X_6	Number of output formats	0.68	0.39
X_7	Characters in data base	0.31	0.73
X_8	Number of data base record types	0.24	0.85

Dependent Variables			
Symbol	Name	Factor I	Factor II
Y_1	Man-months of development effort	0.40	0.63
Y_3	Number of program maintenance personnel	0.51	0.77
Y_4	Number of operations personnel	0.74	0.52
Y_5	Dollars per month of hardware cost for application production	0.93	0.22
Y_6	Dollars per month of hardware cost for program maintenance	0.91	0.17

Note: High factor loadings are underlined.

F. Findings and Conclusions

1. Findings

Workload descriptors were refined by defining them more precisely, by eliminating nonsignificant descriptors, and by combining significant descriptors. The small set of workload descriptors consists of the following:

- o Characters per month of input volume
- o Number of input data fields
- o Characters per month of output volume
- o Number of output formats
- o Characters in data base

Workload descriptors were shown to have breadth of application; that is, they can be applied to a wide variety of systems. Workload descriptors were not conclusively proven to exhibit inclusiveness and exclusiveness; that is, they are similar for some similar systems and dissimilar for some dissimilar systems, but not always consistently so.

Relationships between workload descriptors and costs were determined, and cost estimation equations were derived for the following cost variables:

- o Man-months of development effort
- o Number of program maintenance personnel
- o Number of operations personnel
- o Dollars per month of hardware cost for application production
- o Dollars per month of hardware cost for program maintenance

The relationship between workload descriptors and elapsed development time were not derivable from the 18-ADPS sample, but may be obtained as a function of the number of man-months of development effort.

The estimating relationships derived from the 18-ADPS sample displayed wide prediction intervals. With an increase in sample size from 18 to N, the $\log_{10}$ interval width would be decreased at least by a factor of $\sqrt{18/N}$ if everything else remained the same or improved.

2. Conclusion

Relationships do exist between workload descriptors and costs.

IV. EXPERIENCE HANDBOOK DEVELOPMENT

This section discusses the objectives of and the activities leading to the development of the ADP Experience Handbook (Pilot Version).

A. Objectives

The Phase II goals for development of the Experience Handbook were as follows:

- Develop a method to facilitate the solution of cost estimation equations
- Write 18 ADPS macrodescriptions
- Develop indexing schemes for finding portions of the macrodescriptions relevant to a proposed ADPS based on attributes of the proposed ADPS
- Organize solutions to cost estimation equations, macrodescriptions, and indexes into a usable handbook

To achieve these objectives, the following tasks were performed:

- Development of cost estimation graphs
- Development of "total" system descriptions of fixed format consisting of seven pages of highly distilled information
- Development of indexing methods and procedures for discriminating retrieval of relevant data
- Organization of a handbook in easily usable form
- Development of procedures for use of the handbook
- Establishment of a glossary of terms
- Preparation of a primer as a more detailed example of how the handbook is used

Each of these tasks is described in the subsequent sections, followed by a section on findings and conclusions.

B. Cost Estimation Graphs

1. Rationale for Development of Graphs

An important step in judging proposals for new automation is to conduct a cost analysis. The planned costs for development and operation of an ADPS would be compared to predicted costs that were determined by use of cost estimation equations. Phase II cost estimation equations are given in Table 6, along with equations for determining their 80 percent prediction intervals.

The solution of these equations would require the transformation of the X variables by use of logarithms and finally the retransformation of the results by obtaining the antilogarithms. In addition, a square root must be computed during the computation of the prediction interval. The user of these equations must therefore be conversant in the use of logarithms. The calculation itself is quite a burdensome chore subject to clerical errors. As a result, several methods were investigated to provide means for aiding the proposal evaluator in this task. Two of these methods are described in the following paragraphs.

2. Graphical Aids to Computation

a. Nomographs

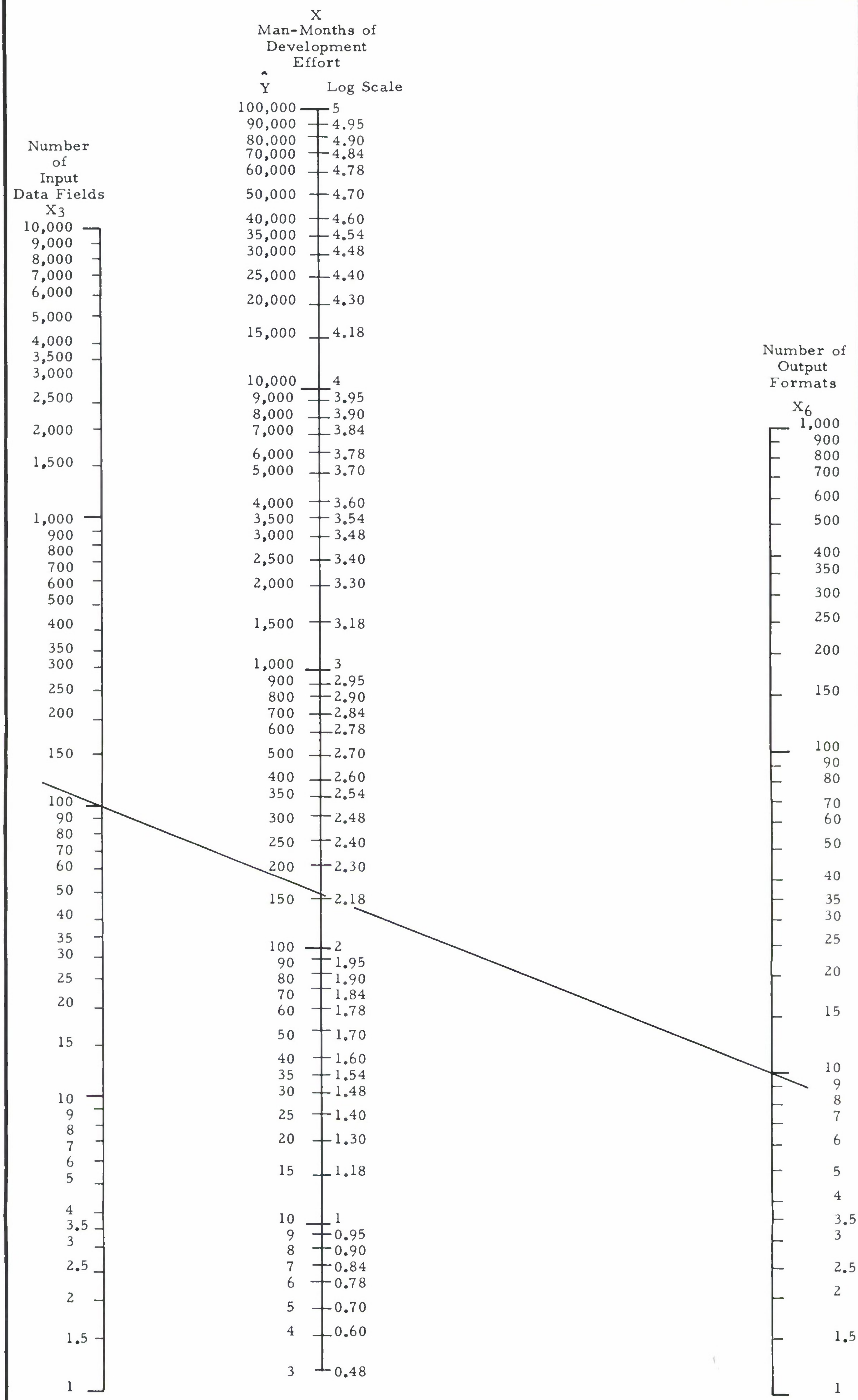
A set of nomographs was developed to aid in computing the predicted cost values and the prediction intervals. An example of the nomographs for the solution of one set of equations is given in Figures 11 and 12. The use of these nomographs to obtain the predicted value and prediction interval would entail the following steps:

- Locating six points on the scales
- Drawing four straight lines
- Reading two values
- Manually adding and subtracting two values
- Locating three additional points on a scale
- Reading three additional values

This method substantially reduced the computational effort but was still rather cumbersome. Therefore, the search for simpler methods was continued.

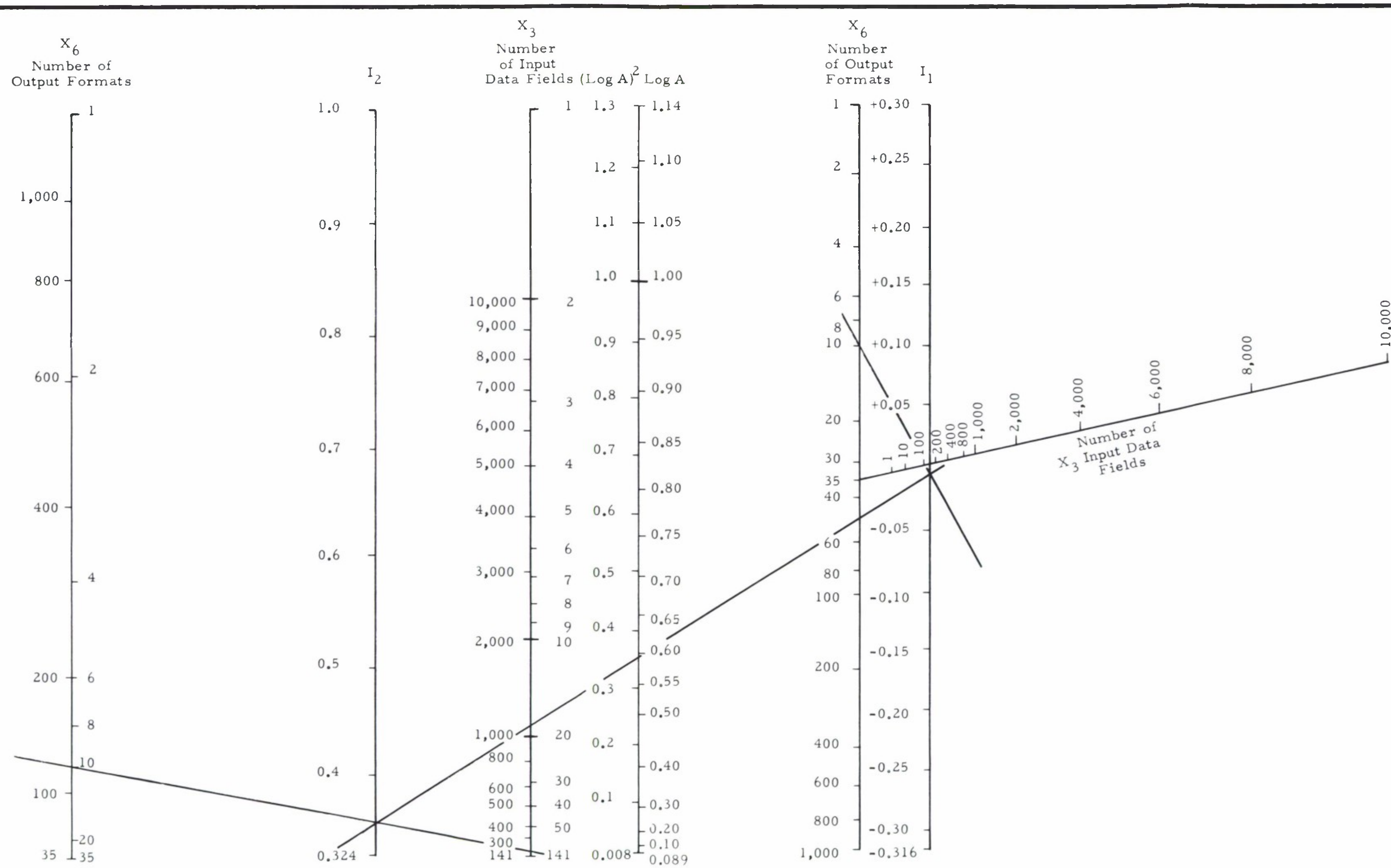
b. Iso-Graphs

A set of iso-graphs was then developed in an attempt to further simplify the computation task. An example of the iso-graphs



$$\log_{10}\left(\begin{matrix}\hat{Y} \\ \text{Predicted} \\ \text{Man-Months of} \\ \text{Development} \\ \text{Effort}\end{matrix}\right) = 0.426 + 0.692 \log_{10}\left(\begin{matrix}X_3 \\ \text{Number of} \\ \text{Input Data} \\ \text{Fields}\end{matrix}\right) + 0.376 \log_{10}\left(\begin{matrix}X_6 \\ \text{Number of} \\ \text{Output} \\ \text{Formats}\end{matrix}\right)$$

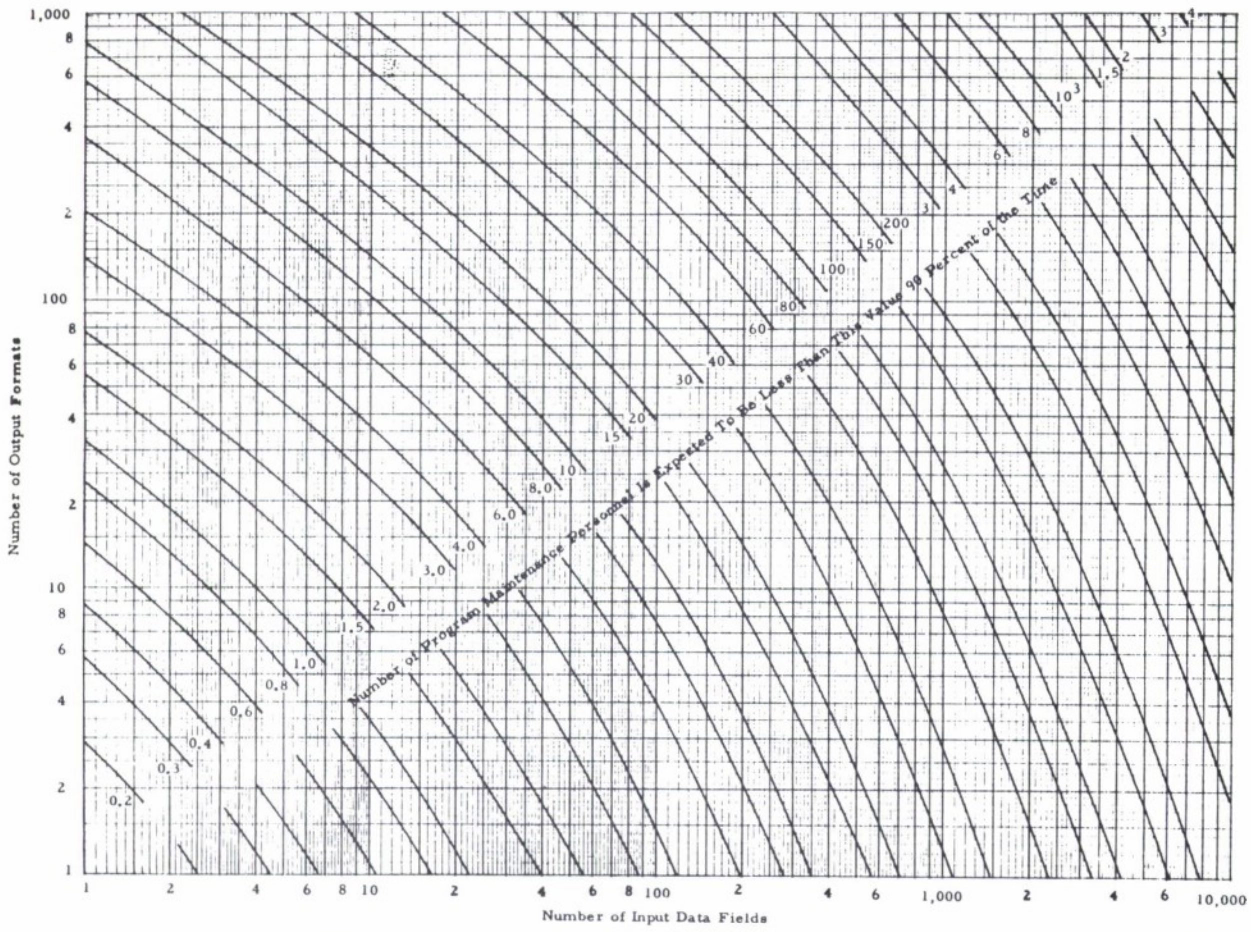
FIGURE 11 - PREDICTION EQUATION NOMOGRAPH FOR MAN-MONTHS OF DEVELOPMENT EFFORT

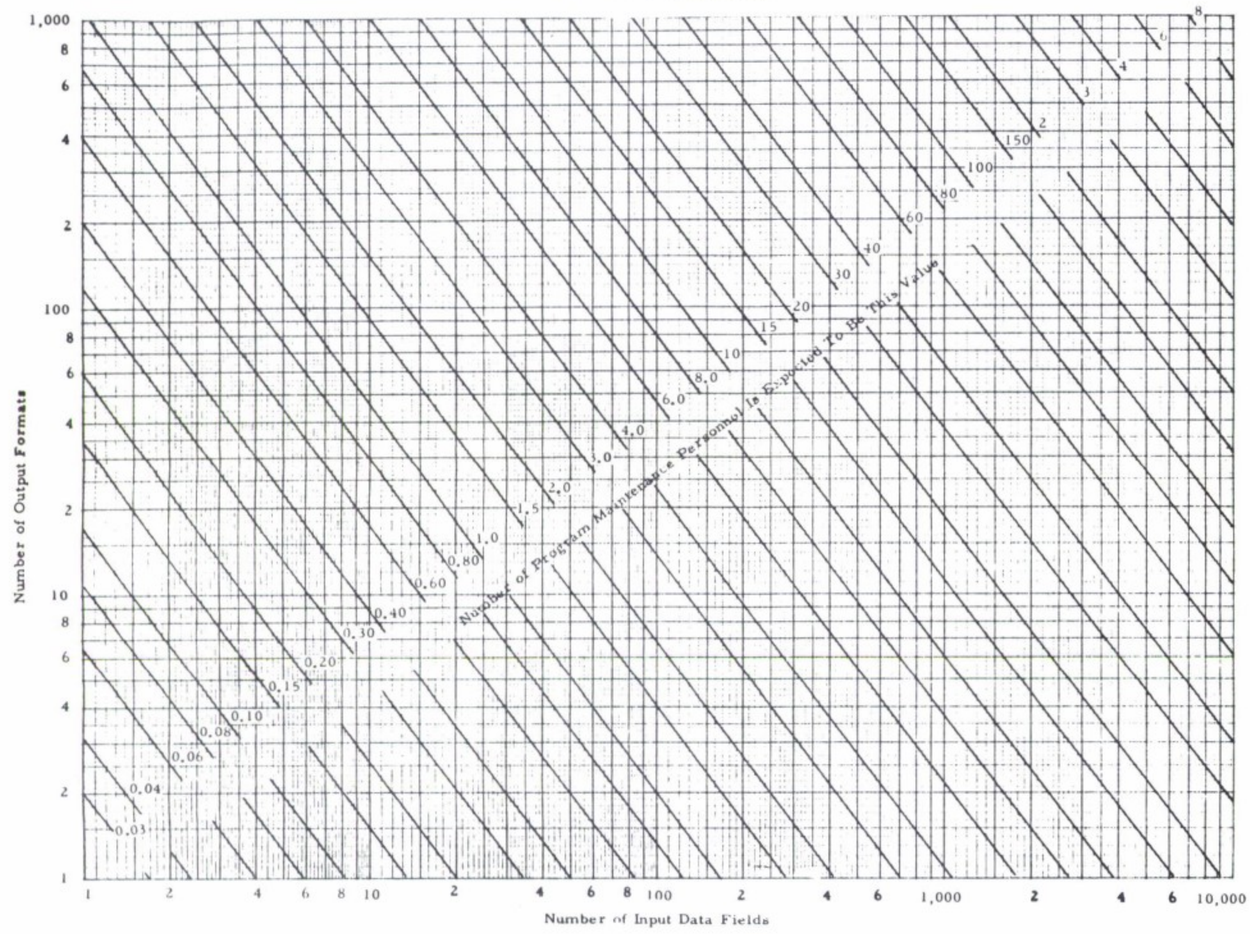


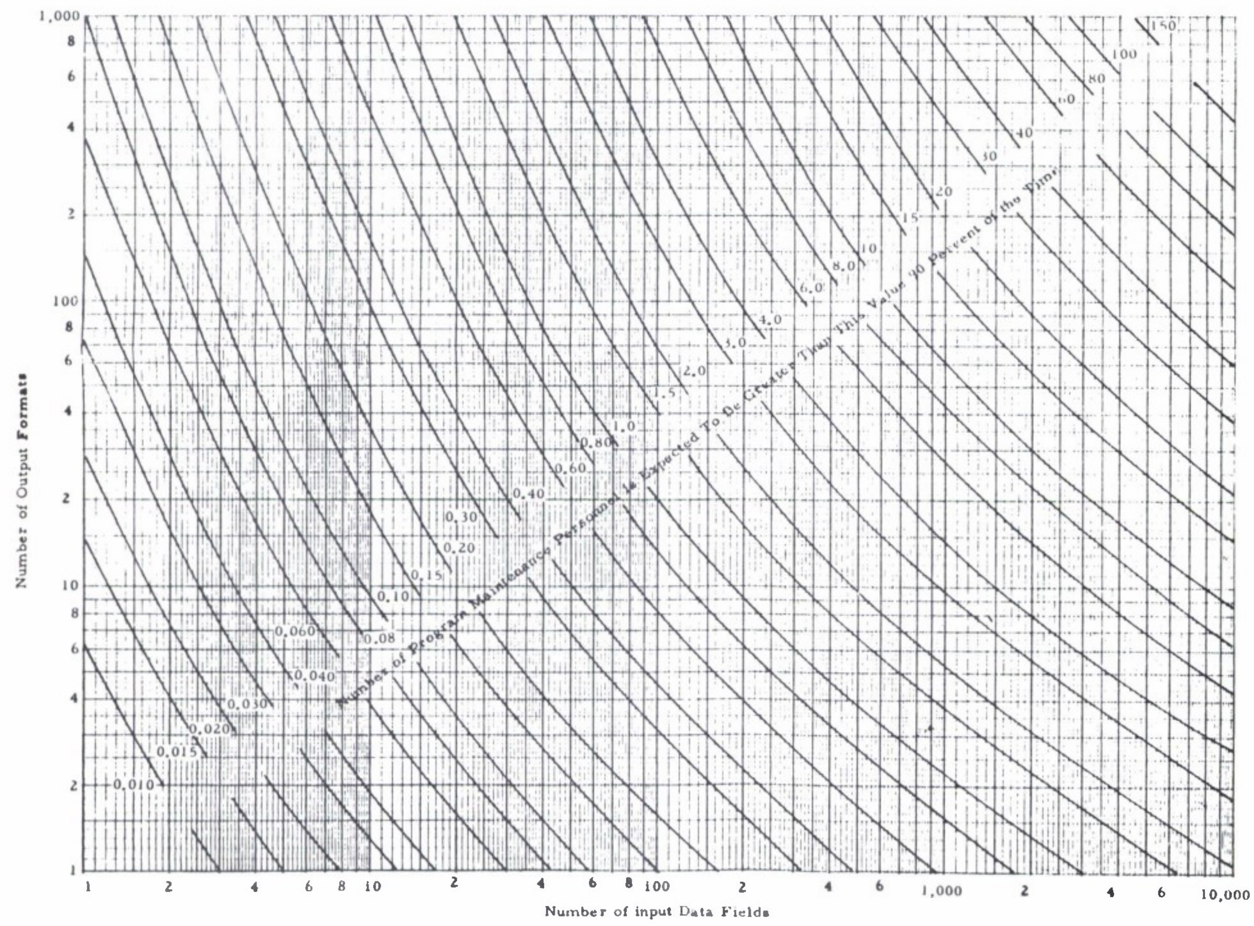
$$\text{Log } A = 0.554\left[1.055 + 0.321(\text{Log}_{10}X_3-2.149)^2 + 0.311(\text{Log}_{10}X_6-1.544)^2 - 0.360(\text{Log}_{10}X_3-2.149)(\text{Log}_{10}X_6-1.544)\right]^{1/2}$$

Where X_3 = Number of input data fields
X_6 = Number of output formats

FIGURE 12 - PARTIAL EIGHTY PERCENT PREDICTION INTERVAL NOMOGRAPH FOR MAN-MONTHS OF DEVELOPMENT EFFORT







Cost Estimating Procedure for Number of Program Maintenance Personnel

1. Find the value of Number of Input Data Fields for the proposed ADPS on the horizontal scale of any one of the three iso-graphs.
2. Draw a vertical line through all three iso-graphs at the value established in Step 1.
3. Find the value of Number of Output Formats for the proposed ADPS on the vertical scale of each of the three iso-graphs.
4. Draw a horizontal line on all three iso-graphs through the values established in Step 3.
5. On the top iso-graph, determine the value that Number of Program Maintenance Personnel is expected to be less than, 90 percent of the time, by logarithmically interpolating the intersection point of the vertical (Step 2) and horizontal (Step 4) lines between adjacent iso-lines.
6. On the center iso-graph, determine the value that Number of Program Maintenance Personnel is expected to be, by logarithmically interpolating the intersection point of the vertical (Step 2) and horizontal (Step 4) lines between adjacent iso-lines.
7. On the bottom iso-graph, determine the value that Number of Program Maintenance Personnel is expected to be greater than, 90 percent of the time, by logarithmically interpolating the intersection point of the vertical (Step 2) and horizontal (Step 4) lines between adjacent iso-lines.

FIGURE 13 - EXAMPLE OF COST ESTIMATING ISO-GRAPHS FOR NUMBER OF PROGRAM MAINTENANCE PERSONNEL

for one set of equations is given in Figure 13. The use of iso-graphs to obtain the predicted value and the prediction interval would entail:

1. Locating four points on the scales.

2. Drawing four straight lines.

3. Reading three values.

This method, which requires considerably less effort than using the nomographs, is the method selected for use in the ADP Experience Handbook. Workload descriptors are used to retrieve the cost estimates. Instructions for using the iso-graphs can be found on the charts. For more detailed descriptions and instructions and a complete set of these iso-graphs, see the ADP Experience Handbook.

C. <u>System Descriptions</u>

A seven-page system description was developed for each system, using data collected in this phase. The descriptions, which are included in the Air Force ADP Experience Handbook (Pilot Version), are highly formatted and standardized to provide rapid cognition of system problems and attributes and to enhance cross-system comparisons. Each system description presents a total system picture which comprises the following 21 sections:

- System
- Data System Designator
- Data Collection Date
- Location
- Function
- Organization
- History
- Schedule
- Description
- Workload
- Hardware
- Software
- Application Program Development

- File Conversion
- Documentation
- Personnel
- Operations
- Application Program Maintenance
- Benefits
- Cost Factors
- Future Plans

Information from the system descriptions is retrieved through the use of one or more of the indexing schemes (see subsection IV. D). The use of an index will normally retrieve only specific sections of a system description. Because complete understanding of a specific section may entail examining other sections, it was preferable to organize system descriptions by system, with all information on a system grouped together.

The information contained in a system description can be broken into the two broad categories of descriptive or explanative information. Descriptive information defines the magnitude and nature of the subject system, while explanative information gives reason for occurrence of problems and attributes of the system. Normally, in retrieving information from a system description for purposes of evaluation, one is primarily interested in the explanative information as clarified and put in context by the descriptive information. Purely descriptive data pertaining to cost factors are obtained through the use of the cost estimating iso-graphs.

D. Indexes and Use

Indexes were developed for use in retrieving information from the system descriptions. The twelve indexes fall into two categories: (1) continuous workload descriptor indexes and (2) discrete system attribute indexes.

1. Continuous Workload Descriptor Indexes

The Development Experience Index and the Operations Experience Index are both based on the use of workload descriptors for retrieval of relevant experience data. The Development Experience Index uses workload descriptors that are causally related to development cost and problems. The following workload descriptors were found to be most suitable for retrieval of development experience:

- Number of input transaction types
- Number of input data fields
- Number of output formats
- Number of data base record types

The workload descriptors provided measures of the size of development effort. Therefore, development experience data retrieved by this index would relate to problems caused by the size of the development effort. An example of the type of experience data retrieved by this index could be the necessity of establishing formal lines of communication between analysts and programmers for systems of the size being evaluated.

The Operations Experience Index scheme was quite similar in structure to the Development Experience Index. The workload descriptors that are causally related to operations cost and problems were the following:

- Characters per month of input volume
- Characters per month of output volume
- Characters in data base

These workload descriptors provided a measure of the size of operations effort. Therefore, operations experience data retrieved by this index would relate to problems caused by the size of the operations.

The construction of the Development Experience Index is described in the following paragraphs. A similar construction was used for the Operations Experience Index. The complete range of each of the four workload descriptors used in the Development Experience Index was represented by separate parallel logarithmic scales. See Figure 14 for an example of these scales. On each scale, the sampled value of the workload descriptor for each of the 18 systems was marked and labeled. Logarithmic scales were chosen for the same reasons as for the transformation of the original variable values (see subsection III. B. 5).

To find relevant systems, a transparent index card with slides for operations indexing and development indexing was constructed. The slides are marked with tolerance bands, designated by ranking numbers, that represent a fixed percentage difference from the proposed value of a workload descriptor. The tolerance bands and their weights are as follows:

Tolerance From Proposed Value	Rank
±15 percent	3
±30 percent	2
±45 percent	1

The tolerance bands were determined empirically by examining systems selected with different tolerance bands. The different width of tolerance bands for the operations slide and development slide results from the different logarithmic scales used with the slides.

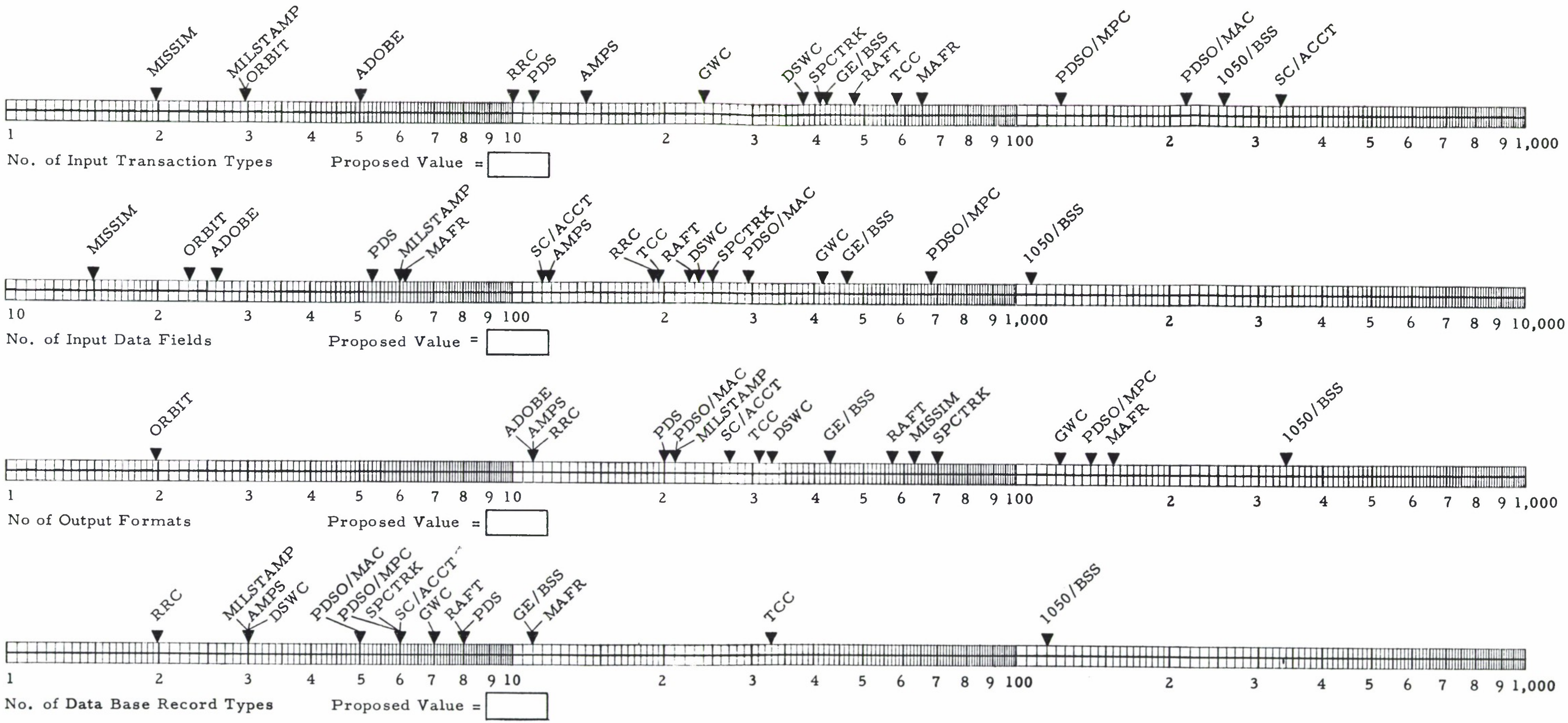
The development slide is used by centering it over the proposed value of Number of Input Transaction Types and entering in the ranking table the rank of all systems bounded by the tolerance bands on the slide. This is repeated for the other workload descriptors. The total rank for a system is determined by adding that system's rank for each individual workload descriptor. The relevancy of systems is in order of total rank, with the system of highest total rank having greatest relevancy.

2. Discrete Attribute Indexes

Ten different indexes were available for retrieving relevant data based on discrete system attributes:

- Functional Area Index
- Decentralized Operations Index
- Multiple Application Index
- Programming Language Index
- Processing Type Index
- File Conversion Index
- Direct Access Storage Index
- Computer Cost Index
- Computer Index
- Security Index

With these indexes, all 18 systems were classified into three or more categories based on the attribute defined by the index name. The attribute of the proposed system was used to isolate all systems in the same



Use of the Development Experience Index

The development experience index is used to retrieve relevant development experience from the system descriptions. Retrieval is based on the following workload descriptors:

- Number of Input Transaction Types
- Number of Input Data Fields
- Number of Output Formats
- Number of Data Base Record Types

Proposed values of these workload descriptors must be known for the ADPS under evaluation. The type of development experience and problems retrieved by this index will be related to the magnitude of the development effort. Sections of system descriptions which will be of greatest interest include:

Application Program Development
Application Program Maintenance
Schedule

Other sections of interest include:

Organization
History
Software
File Conversion
Documentation
Personnel
Benefits
Cost Factors
Description

Procedure

1. Enter the proposed values for No. of Input Transaction Types, No. of Input Data Fields, No. of Output Formats, and No. of Data Base Record Types in box below the corresponding scale.
2. Remove the index card from the pocket in the Air Force ADP Experience Handbook (Pilot Version) and position the center arrow of the Development Slide at the proposed value on the No. of Input Transaction Types scale.
3. For all systems bounded by the Development Slide, enter the number from the tolerance band in the No. of Input Transaction Types row of the Ranking Table beneath the corresponding system name.
4. Repeat steps 2 and 3 for No. of Input Data Fields, and No. of Output Formats.
5. Repeat steps 2 and 3 for No. of Data Base Record Types if the proposed value is not zero. If the proposed No. of Data Base Record Types is zero, enter the number 3 in the Data Base row of the Ranking Table beneath ADOBE, MISSIM, and ORBIT.
6. Enter the Total Rank in the bottom row of the Ranking Table. Total Rank is computed by adding the numeric entries in each each column of the Ranking Table.
7. The system with the largest Total Rank is the most relevant system in developmental aspects to the proposed system. Relevancy of other systems is in order of Total Rank. Systems with Total Rank equal to or greater than 7 are highly relevant to the proposed automation. Systems with Total Rank less than 7 but greater than 3 have less relevance but may still be used, while developmental experience data from systems with Total Rank less than or equal to 3 should not be used.

Ranking Table

	ADOBE	AMPS	DSWC	GE/BSS	GWC	MAFR	MILSTAMP	MISSIM	ORBIT	PDS	PDSO/MAC	PDSO/MPC	RAFT	RRC	SC/ACCT	SPCTRK	TCC	1050/BSS
Page No.																		
No. of Input Trans. Type Rank																		
No. of Input Data Fields Rank																		
No. of Output Formats Rank																		
No. of Data Base Record Types Rank																		
Total Rank																		

FIGURE 14 - WORKSHEET FOR DEVELOPMENT EXPERIENCE INDEX

category through the use of an index table. The isolated systems will contain relevant experience data in the area of the category retrieved. Thus, if the proposed system used COBOL, all sampled systems in the COBOL category would be retrieved from the programming language attribute index.

E. ADP Experience Handbook Integration and Use

Following the development of cost estimation iso-graphs, system descriptions, and indexing methods, the handbook sections were organized and integrated into an easily usable form. Instructions for the use of the iso-graphs in cost prediction are given in one section, and an index is provided for retrieving relevant information from the system descriptions. Terms used in the handbook are defined in a glossary.

The integration of these sections resulted in a self-sufficient handbook. A Primer for the use of the handbook was developed to provide an elementary text that can be used to train potential users of the handbook. The Primer contains instructions for submission of ADPS proposals, a sample ADPS proposal, and an evaluation of the sample ADPS proposal.

The Air Force ADP Experience Handbook (Pilot Version) and the Primer for the Air Force ADP Experience Handbook comprise two separate volumes; refer to PRC documents R-930 and R-931, respectively.

F. Findings and Conclusions

1. Findings

- Iso-graphs are an effective aid to the manual solution of cost estimation equations.
- Macrodescriptions of ADP systems can be written.
- The indexing attributes are as follows:

  Workload descriptors (small sensitive set of seven)

  Functional area

  Decentralized operations

  Multiple applications

  Programming language

  Processing type

File conversion type

Direct access storage capacity

Computer cost

Computer make and model

Security requirements

- o Iso-graphs, macrodescriptions, and indexes are organizable into a usable handbook.

2. Conclusions

The ADP Experience Handbook will be a useful tool for personnel concerned with management of ADP with the Air Force.

V. PHASE III PLANNING

The purpose of this section is to briefly review the planning activity that preceded the actual development of the operational concept and development plan for Phase III. The activities of data collection, data analysis, and experience handbook development to the midpoint of Phase II have led to preliminary conclusions. When integrated, these conclusions indicated that Phase III will be desirable to the Air Force if it is feasible and cost effective. Hence, the Phase III operational concept, development plan, and the analysis of costs and benefits were developed and are covered in detail in Volume III of this final report.

The primary goal of the Phase III planning activity was to establish a concept for collecting, editing, reducing, and using data processing data at HQ USAF. It was therefore necessary to learn more about who used such data and for what purpose, who reported such data and in what detail, what was the general content of an ADPS proposal, and what evaluation processes took place, etc.

Accordingly, project staff members visited the Air Staff on several occasions to collect such information, principally to determine in detail Data Automation Proposal (DAP) evaluation procedures and policies as well as ADP experience reporting procedures. After lengthy discussions with members of AFADA and AFSPD and after analysis of all appropriate regulations, manuals, and operating instructions, a general picture of ADPS proposal procedures and reporting procedures was established. These findings are summarized in Appendixes F and G.

It became evident during this task that Data Automation Proposals are only one of several types of ADPS proposals that must be judged by HQ USAF and, further, that there are several places in HQ USAF that the judging takes place. In addition to DAP's, the following documents can propose systems: Required Operational Capability (ROC); Requirements Action Directive (RAD); Advance Communications -- Electronic Requirements Plan (ACERP); Communications -- Electronics Implementation Plan (CEIP); Program Change Proposal (PCP); Proposed System Package Plan (PSPP); System Package Plan (SPP); Preliminary Technical Development Plan (PTDP); and others. Essentially, all parts of the Air Staff can become involved in the evaluation of a proposal; however, the designated offices of primary responsibility (OPR's) include the Director of Data Automation (AFADA), the Director of Production (AFSPD), the Director of Maintenance Engineering (AFSME), the Director of Operational Requirements and Development Plans (AFRDQ), and the Assistant for Research and Development Programming (AFRRP).

The scope of the Phase III planning activity was broadened to ensure that the concept established for implementation in Phase III of the project met the ADP management needs of all parts of the Air Staff. Other Air Staff ADP management functions include efficient utilization of ADP assets, prosecution of ADP standards programs, control of on-going ADP developments and operational systems, forecasting of ADP budgets, and performance of special studies of Air Force ADP.

The preliminary conclusion of Phase III planning is that the central feature of Phase III should be an Air Force ADP management information system capable of systematically collecting, editing, storing, retrieving, and putting to use experience and asset data from all Air Force ADP systems and data processing installations. Although the principal need for the system is at HQ USAF, other Air Force organizations could make use of it (e. g., SPO's).

APPENDIX A

AIR FORCE AND CIVILIAN PERSONNEL CONTACTED DURING DATA COLLECTION

Appendix A provides a list of the personnel from whom data were collected at the 18 installations interviewed. Names of PRC interviewers, location of the installation, and dates of interviews are also provided.

Dates	System	Location	PRC Interviewers	Primary Contacts
14-21 Mar 1966	MPC PDSO	Randolph AFB San Antonio, Texas	G. Beckwith A. Gradwohl L. Wimpey S. Wong	Col. J. Monahan Col. W. Best Col. L. Reed Lt. Col. W. Fortin Lt. Col. R. Glass Lt. Col. J. Harrop Maj. H. Axmacher Maj. P. Bundick Maj. H. Styron Maj. C. Weltz Maj. E. Willard Capt. H. Arthur Capt. R. Jackson SMSgt. C. Grindler SMSgt. H. McGuire TSgt. C. Christy TSgt. G. Poole SSgt. C. Ellis SSgt. K. Nichols Mr. J. Arnold Mr. L. Chambless Mr. B. Dix Mr. D. Essenpreis Mr. E. Foster Mr. C. Jagge Mr. L. Maddux Mr. M. Volluz
11-14 Apr 1966	Missile Simulation System	Eglin AFB Valparaiso, Florida	A. Gradwohl A. Sepan	Lt. Col. J. McClellan Sgt. C. Sessons Mr. F. Bryan Mr. C. Davy Mrs. C. Elder

Dates	System	Location	PRC Interviewers	Primary Contacts
				Mr. G. Fikes Mr. D. Harrison Mr. C. Kidd Mr. W. Mixon Mr. W. Spier Mr. J. Thoreen Mr. W. Thweatt Mr. G. Weekly Mr. W. Willslette, Jr. Mr. H. Wohlin Mr. K. Woolsey
11-15 Apr 1966	RAFT	Randolph AFB San Antonio, Texas	L. Roseman S. Wong	Maj. D. Bye Mr. R. Boenig Mr. O. Henthorn Mr. C. Keyes Mr. J. Oates Mr. W. Thompson Mr. C. Trefzer
11-15 Apr 1966	GE 225 Base Inventory	Scott AFB St. Louis, Missouri	G. Beckwith C. Remstad	Lt. Col. J. Mumma Lt. Col. W. Spann Capt. J. Gates SMSgt. W. Bleakney SMSgt. R. Votter TSgt. J. Brewer TSgt. H. Thompson Mr. W. Brinkman Mr. R. Land Mr. A. Kiefer Mr. H. Sutherland

Dates	System	Location	PRC Interviewers	Primary Contacts
11-15 Apr 1966	Priority Distribution System	Wright-Patterson AFB Dayton, Ohio	D. Kiernan L. Wimpey	Col. C. Spencer Mr. J. Brown Mr. H. Hartsock Mr. V. Jobe Mr. C. Lovett
18-22 Apr 1966		McClellan AFB Sacramento, California		Col. J. Shinners Mr. D. Cunningham Mr. H. Davis Mr. L. Gearhart Mr. D. Halford Mr. J. Kurtz Mr. S. Lambert Mr. L. Taylor
14-19 Apr 1966	MAC PDSO	Randolph AFB San Antonio, Texas	L. Roseman S. Wong	Col. J. Evans Col. J. Monahan Lt. Col. R. Mebane Lt. Col. R. Murray Maj. P. Phillips Capt. A. Manion Lt. R. Householder MSgt. L. Zeitner SSgt. J. Brummer Mr. R. Roe Mr. R. Spurlock Mr. C. Trefzer Mr. B. Wiley
15-20 Apr 1966	SPACETRACK	Ent AFB Colorado Springs, Colorado	A. Gradwohl A. Sepan	Col. W. Arnold Col. H. R. Ebbeler Col. E. Gravette Lt. Col. J. Gregor Lt. Col. Szczulkowski

Dates	System	Location	PRC Interviewers	Primary Contacts
				Maj. T. Birge Capt. W. Bolmanski Capt. G. Ellefritz Capt. Gingrich Capt. H. VanDerryt Capt. J. Wilde Lt. R. Morrison Lt. G. Roeder SSgt. D. Armstrong Mr. R. Ferguson Mr. R. Jackson
18-22 Apr 1966	MILSTAMP (Central)	McClellan AFB Sacramento, California	D. Kiernan L. Wimpey	Col. J. Shinners Mr. A. Carlson Mr. H. Davis Mr. B. Scarborough Mr. W. Speer
18-22 Apr 1966	Global Weather Central	Offutt AFB Omaha, Nebraska	G. Beckwith C. Remstad	Col. R. Johnston Col. R. Osborn Lt. Col. L. Jacobs Lt. Col. W. Miller Maj. B. Aldridge Maj. C. Cook Maj. A. Gargailo Maj. Johnston Maj. M. Warner Capt. G. Delong Capt. R. Leclair Mr. A. Gulliver Mr. B. Lawrence

Dates	System	Location	PRC Interviewers	Primary Contacts
9-10 May 1966	ORBIT Determination	A. F. Cambridge Research Laboratories L. G. Hanscom Field, Bedford, Massachusetts	A. Sepan S. Wong	Mr. J. Cooney Miss E. Cronin Mr. R. Gosselin Mrs. I. Hussey
9-12 May 1966	AMPS	AFAFC Denver, Colorado	G. Beckwith L. Wimpey	Col. B. Blair Lt. Col. R. Strada Maj. Grey Lt. Derby TSgt. V. Rhode Mr. W. Davis Mr. K. Grant Mr. J. Keefe Mr. J. Thomas
9-13 May 1966	AFSC Accounting	Andrews AFB Washington, D.C.	D. Kiernan C. Remstad	Col. R. O'Hara Mr. H. Ichoff Mr. Priestly Mr. J. Reino Mrs. J. Ruddel
11-13 May 1966	TAC Command and Control	Langley AFB Hampton, Virginia	A. Sepan S. Wong	Col. F. Cloninger Lt. Col. Biggs Lt. Col. Frank Lt. Col. Hull Maj. Major Maj. Rhodes Capt. Anunson Capt. Stallings SMSgt. Northrup Mr. V. Cook Mr. D. Rohrer

Dates	System	Location	PRC Interviewers	Primary Contacts
13-19 May 1966	MAFR	AFAFC Denver, Colorado	G. Beckwith L. Wimpey	Col. B. Blair Lt. Col. R. Strada Mr. E. Christesen Mr. R. Gorin Mr. N. Kent Mr. E. Krischel Mr. M. Mallory Mr. J. Musser Mr. G. Muldowney Mrs. M. Powell Mr. M. Taylor Mr. H. Wise
16-19 May 1966	Adobe	Edwards AFB Edwards, California	A. Gradwohl L. Roseman	Mr. C. Graham Mr. C. Herron Mr. D. Jared Mr. W. Mayer Mr. R. Noblin Mr. D. Shoemaker Mr. G. Wilson
16-20 May 1966	1050 Base Supply System	Bolling AFB Washington, D.C.	D. Kiernan C. Remstad	Col. Cronemiller Col. VanCleef Maj. Bartlett Capt. McConnel Sgt. Haycock
5-11 July 1966	Data System Workload	Wright-Patterson AFB Dayton, Ohio Kelly AFB (SAAMA) San Antonio, Texas	C. Remstad A. Sepan	Col. C. F. Spencer Maj. D. J. Nielsen Capt. J. N. Becker Mr. E. Davlin (SAAMA) Mr. C. E. Leaverett (SAAMA) Mr. R. L. Maeierhofer (SAAMA)

Dates	System	Location	PRC Interviewers	Primary Contacts
				Mr. M. Mitchell Mr. R. Plummer Mr. D. Pue (SAAMA) Mr. R. Recny Mr. P. Rivera (SAAMA) Mr. A. G. Sinclair (SAAMA) Mr. C. A. Stubbs (SAAMA)
5-11 July 1966	Repair Requirement Computation System	Wright-Patterson AFB Dayton, Ohio Tinker AFB (OCAMA) Oklahoma City, Oklahoma Kelly AFB (SAAMA) San Antonio, Texas	C. Remstad A. Sepan	Mr. K. Collier (OCAMA) Mr. W. E. Cooper (OCAMA) Mr. J. S. Husser Mr. J. N. Kemp (OCAMA) Mr. R. Reid (OCAMA) Mr. P. Rivera (SAAMA) Mr. J. Stafford Mr. C. A. Stubbs (SAAMA) Mr. Temple (SAAMA)

APPENDIX B

DATA COLLECTION QUESTIONNAIRE

Appendix B consists of the data collection questionnaire as revised after the pilot data collection effort on the Military Personnel Center Personnel Data System--Officer (MPC PDSO) at Randolph Air Force Base.

Parts A, B, C, D, and E comprise the questionnaire proper, and Part F (Summary Sheet for Numerical Data) is a summary of reduced data for the statistical analysis.

QUESTIONNAIRE OUTLINE

A. General Description

B. Proposal

   1. Content
   2. Preparation

C. Development

   1. Technical Approach
   2. Management Approach
   3. Schedule
   4. Manpower
   5. Hardware
   6. Software Supplied by Others
   7. System Design
   8. Programming
   9. File Conversion

D. Operations

   1. Organization
   2. Manpower
   3. Workload
   4. Scheduling
   5. Utilization
   6. Program Maintenance
   7. Hardware
   8. Support Programming
   9. Facility

E. Future Plans

F. Summary Sheet for Numerical Data

A. GENERAL DESCRIPTION

1. Obtain an organization chart showing position of both ADPS and prime user(s) within the Air Force.
2. Describe the mission of the ADPS.
3. Describe the mission of the prime user(s).
4. Obtain the DSAP symbols and titles (see RCS 8-AF-E6) for the ADPS.
5. Obtain a simplified functional block diagram and succinctly describe the work the ADPS does.
6. Develop a narrative history of ADPS from inception to present day.
7. What are the consequences of ADPS being down for an extended period of time? (Include backup capability.)
8. Describe any security factors involving ADPS and installation.
9. Describe any problems, consequences of problems, and expected solutions.

B. PROPOSAL

1. <u>Content</u>

Describe the content of the DAP (or other paperwork or briefing) upon which Hq. USAF based approval of ADPS. Include the proposer's conception of future events and activities during development and operations, using the following topics as guidelines:

Development	Operations
Tasks to be performed during development	Organization
Organizational approach	Manpower
Schedule	Workload
Manpower	Files
Hardware	Scheduling
Software supplied by others	Utilization
	Program maintenance
	Hardware
	Systems programming
	Facility

2. Preparation

Describe preparation effort for proposal; for example, who did it, how effort was organized, and pertinent dates.

C. DEVELOPMENT

1. Technical Approach

Describe steps or activities undertaken during development, particularly in contrast to what was proposed.

2. Management Approach

   a. Describe organizational approach to development; for example, tasks assigned to various organizational entities and relationships among analysts-programmers-users-management.
   b. Describe management control methods used during development; for example, PERT, cost control, or progress reports.

3. Schedule

(Use sheet provided.)

4. Manpower

   a. Collect data on personnel buildup during development phase. (Use sheet provided.)
   b. See that personnel data sheets are distributed.

5. Hardware

   a. Describe hardware selection process; for example, the RFQ, number of bidders, and selection criteria.
   b. Describe any unusual installation problems.

6. Software Supplied by Others

(Use sheet provided.)

7. System Design

   a. Was this a pioneering application? If so, why?
   b. What is the system design? Obtain a flow chart.

c. What workload was the system designed to handle? Is it different from the workload at the proposal stage and the present workload?

d. To what extent did the system design change during the development stage?

e. Were there interface problems with other ADP systems?

f. Did earlier automation efforts make this development easier?

g. Describe documentation activities with regard to system specification to both programmers and users, and design changes.

8. <u>Programming</u>

a. In your opinion, was the machine "mature" at the time of system development?

b. Describe documentation activities of programmers with regard to program specifications, detailed flow charts, program changes, and manuals for operators, users, and program maintenance personnel.

c. Acquire a list of all programs in the system and classify them according to the following <u>primary</u> program functions:

(1) Input Edit (input conversion, data edit, error and logic checks)

(2) File Maintenance (file update, extract data)

(3) Report Generation (data edit, print, display)

(4) Merge (sequence ordered sets of data)

(5) Compute (arithmetic)

(6) Sort (sequence unordered sets of data)

(7) Query-File Search (search file for desired items, display items)

(8) Control (job scheduling, priority handling, hardware component assignments)

(9) Support (nonapplication programs)

(Use sheets provided.)

d. Were any unusual programming techniques employed (for example, list processing or dynamic memory allocation)?

e. Describe compilation and checkout activities. Include location and type of machine, use of emulator/simulator, whether shop was open or closed, usual turnaround time, and whether special input data were developed for checkout purposes.

f. Describe system test activities.

g. How many computer hours were required for development of this ADPS? Include compilation, assembly, checkout, and tests.

9. <u>File Conversion</u>

Describe file conversion activities. Include storage media before and after, how "clean" the files were, manual transcription and editing requirements, manpower used, special programming requirements, and computer time used.

D. OPERATIONS

1. <u>Organization</u>

Obtain an organization chart for the personnel operating the ADPS.

2. <u>Manpower</u>

a. How many programmers are in program maintenance?

b. How many people are in computer operations? (Include computer operators, schedulers, production control, tape librarians.)

c. How many people are in EAM operations? (Include input batching, keypunch, tab operators.)

d. See that Operations Personnel Data Sheets are distributed.

3. <u>Workload</u>

Complete workload sheets.

4. <u>Scheduling</u>

a. Is there a master schedule? If yes, block in master schedule diagram. (Use sheet provided.) Include maintenance and program testing.

b. Is a daily schedule prepared? If yes, describe procedure and when.

c. Describe procedures for scheduling monthly, quarterly, yearly, and special reports or runs.

d. Describe effect and impact of system monitor or control program on scheduling function.

e. Describe procedure for handling priorities.

f. Describe any other functions of scheduling.

g. Is shop open or closed?

5. <u>Utilization</u>

a. Obtain enough information to construct a "pie chart" for a typical day's utilization of the machine. Include time for program maintenance, downtime, production, and compilation, or however the installation breaks the time down. The RCS 6-AF-E6 and 8-AF-E6 reports will be useful for this purpose. If these reports do not exist, other reports titled "program run analysis" or "computer usage report" may be obtainable.

b. Obtain the following for the application under study:

	Machine Hours Per Month
Input edit	______
File maintenance	______
Report generation	______
Merge	______
Sort	______
Compute	______
Query	______
Program maintenance	______

c. The people working in operations and the nonproductive machine time must be apportioned to the application under study. If the application under study shares the machine with other applications, obtain enough information to make the apportionment on the basis of machine hours utilized for the application.

6. <u>Program Maintenance</u>

   a. What percent of the people involved in program maintenance were in the original system development?
   b. What percent of program maintenance can be considered to be corrections and what percent improvements?
   c. What type of documentation is produced for program changes?
   d. What is the average turnaround time for checkout work?

7. <u>Hardware</u>

   a. Obtain enough information to diagram the current hardware configuration including manufacturers and model numbers for each component. Use the RCS 6-AF-E6 report, if available.
   b. How has hardware configuration changed since the system was declared operational?
   c. What is monthly rental or original purchase price?
   d. Comment on the reliability of the hardware.

8. <u>Support</u>

   a. Describe current systems programming activities. (Systems programming is the maintenance and development of compilers, assemblers, control programs, and utility routines.)
   b. Describe data storage activities including tape libraries, card libraries, and physical handling methods.
   c. Describe PCAM and keypunch activities.

9. <u>Facility</u>

   a. Is there adequate space for movement of men and materials in the computer room?

   b. Do operators have good visibility of equipment status indicators and peripheral equipment?

   c. Does standby power exist?

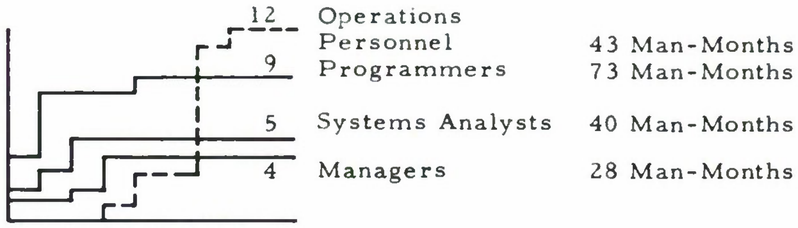
E. FUTURE PLANS

1. What is planned for the future?

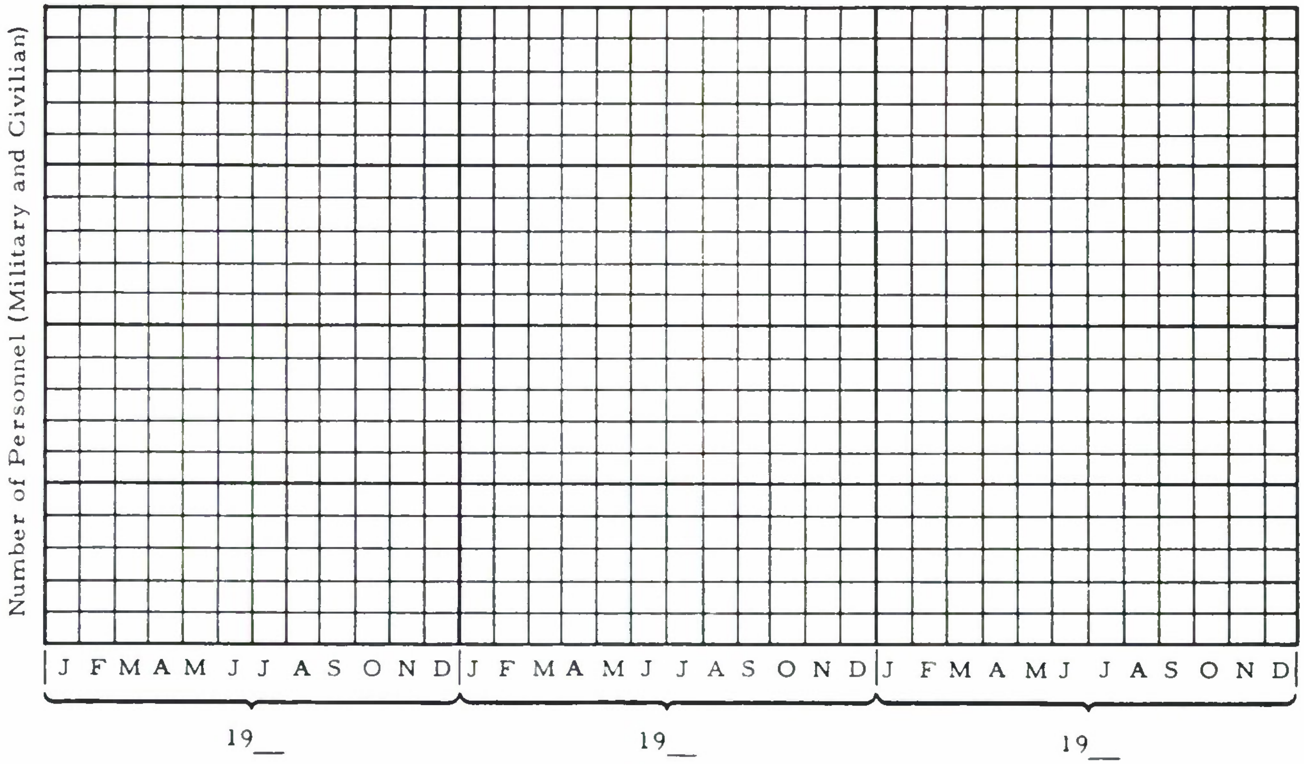
DEVELOPMENT SCHEDULE

Show planned and actual. Include the date the DAP was approved (or work started) and the date the system went operational. Use a task breakdown applicable for the ADPS in question.

19__												19__												19__											
J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D



Example



DEVELOPMENT PERSONNEL BUILDUP BY CATEGORY

DATA SHEET FOR DEVELOPMENT PHASE MANAGERIAL AND LEAD PERSONNEL

(Please fill out your line and pass the sheet on)

Name	Rank or GS No.	Job Title	Education			Experience (Yrs.)		
			High school graduate?	Years college	Degree(s)	In data processing	In using this language	In the field of:

DATA SHEET FOR DEVELOPMENT PHASE PERSONNEL

(Please fill out your line and pass the sheet on)

Name	Rank or GS No.	Job Title	No. of Years Experience	
			In Data Processing	In Field of: ________

SOFTWARE SUPPLIED BY OTHERS

Type of program or subsystem	Supplied by whom?	Initial delivery date	Full effectivity delivery date	Describe any special modifications required for this application	What are the shortcomings of this software?	Describe how this software aided or hindered system development

CHARACTERISTICS OF PROGRAMS

Program Designation	Program Name	Program Function (Sort, Merge Input Edit, etc.)	Developed by Manufacturer, Contractor, or AF	Language in Which Program Was Written	Number of Source Statements	Number of Machine Instructions

DATA SHEET FOR OPERATIONS PHASE MANAGERIAL AND LEAD PERSONNEL

(Please fill out your line and pass the sheet on)

Name	Rank or GS No.	Job Title	Education			Experience (Yrs.)		
			High school graduate?	Years college	Degree(s)	In data processing	In using this language	In the field of: ______

DATA SHEET FOR OPERATIONS PHASE PERSONNEL

(Please fill out your line and pass the sheet on)

Name	Rank or GS No.	Job Title	No. of Years Experience	
			In Data Processing	In Field of: ____________

INPUTS

Description	Source	Batched or direct? (At computer)	Ave. no. of records per mo.	Ave. no. of char. per record	Max. no. of records per mo.	Max. no. of char. per record	No. of unique fields	No. of types of trans-actions	Average freq. of arrival (At com-puter) (Specify units)	Max. freq. of arrival (At com-puter (Specify units)	Ave. % of rec-ords re-jected by input edit

FILES

				If applicable						
Description	Storage medium	Ave. no. of records	Ave. no. of char. per record	Max. no. of records	Max. no. of char. per record	Net growth per month (specify units)	% of fields which have variable length entries	Fre-quency of update	Fre-quency of use	Include in nonre-dundant set?

Note: "Variety" is the number of files in nonredundant set.

OUTPUTS

Description	Recipient	Batched or direct? (At computer)	Ave. no. of units per month	Ave. no. of char. per unit	Max. no. of units per month	Max. no. of char. per unit	No. of diff. types of units (If not all same format)	Response time

MASTER SCHEDULE

	Mon	Tues	Wed	Thurs	Fri	Sat	Sun
0							
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							
23							

F. SUMMARY SHEET FOR NUMERICAL DATA

System____________________ Location________________________ Analyst__________

Date____________

Name of Variable	Must Collect	Collect if Applicable	Collect if Easily Obtainable	Value of Variable or Remarks
Independent variables for				
regression analysis				
Input - Batched				
1. Ave. volume per month	x			
2. Max. volume per month		x		
3. Ave. frequency per hour		x		
4. Max. frequency per hour		x		
5. Variety$_1$ (No. types of				
transactions)		x		
6. Variety$_2$ (No. unique data				
fields)	x			
7. Variety$_3$ (No. unique				
parameter fields)		x		
8. Reliability		x		
Input - Unbatched	x			
9. Ave. volume per month				
10. Max. volume per month		x		
11. Ave. frequency per hour	x			
12. Max. frequency per hour		x		
13. Variety$_1$ (No. types of				
transactions)	x			
14. Variety$_2$ (No. unique data				
fields)		x		
15. Variety$_3$ (No. unique				
parameter fields)		x		
16. Reliability		x		

F. SUMMARY SHEET FOR NUMERICAL DATA (Continued)

System____________________ Location________________________ Analyst__________

Date____________

Name of Variable	Must Collect	Collect if Applicable	Collect if Easily Obtainable	Value of Variable or Remarks
Output - Indirect				
17. Ave. volume per month	x			
18. Max. volume per month		x		
19. Variety (No. report formats)	x			
20. Response time		x		
Output - Direct				
21. Ave. volume per month	x			
22. Max. volume per month		x		
23. Variety (No. report formats)	x			
24. Response time	x			
Data Base				
25. Average size	x			
26. Maximum size		x		
27. Net growth per month	x			
28. % of data updated per month	x			
29. Variety (No. types of				
records)	x			
30. No. items "kept track of"	x			
31. Net growth per month of				
items "kept track of"	x			
Processing Functions - % of				
instructions for:				
32. Input edit	x			
33. File maintenance	x			
34. Report generation	x			
35. Merge	x			

F. SUMMARY SHEET FOR NUMERICAL DATA
(Continued)

System__________________ Location______________________ Analyst________

Date__________

Name of Variable	Must Collect	Collect if Applicable	Collect if Easily Obtainable	Value of Variable or Remarks
Processing Functions (Cont'd)				
36. Sort	x			
37. Compute	x			
38. Query	x			
Personnel - Development				
Managers				
39. Ave. ranks/GS	x			
40. Ave. years college	x			
41. Ave. years in ADP	x			
42. Ave. years in functional area	x			
Analysts				
43. Ave. rank/GS	x			
44. Ave. years college	x			
45. Ave. years in ADP	x			
46. Ave. years in functional area	x			
Programmers				
47. Ave. rank/GS	x			
48. Ave. years college	x			
49. Ave. years in ADP	x			
50. Ave. years with language	x			
51. Ave. years in functional area	x			
Personnel - Operations				
52. Ave. rank/GS	x			
53. Ave. years college	x			
54. Ave. years in ADP	x			
55. Ave. years in functional area	x			

F. SUMMARY SHEET FOR NUMERICAL DATA
(Continued)

System____________________ Location________________________ Analyst_________

Date___________

Name of Variable	Must Collect	Collect if Applicable	Collect if Easily Obtainable	Value of Variable or Remarks
Hardware				
56. Date of first delivery (from				
Adams Associates)[1]	x			
57. Planning information (any				
input, output, data base				
or processing function				
variables stated at time				
of proposal)			x	
Dependent variables for				
regression analysis				
Personnel cost				
101. Man-months for design and				
implementation	x			
102. No. people in program de-				
velopment and				
maintenance	x			
103. No. people in computer				
operations	x			
104. Development time	x			
Planned and actual dates for:				
105. Hardware installation		x		
106. Compiler/assembler				
operating		x		
107. Executive operating		x		
108. Program test system				
operating		x		

[1] Must be combined with other variables in regression analysis.

F. SUMMARY SHEET FOR NUMERICAL DATA
(Continued)

System____________________ Location________________________ Analyst_________

Date____________

Name of Variable	Must Collect	Collect if Applicable	Collect if Easily Obtainable	Value of Variable or Remarks
109. Application operational		x		
Hardware cost - installation as				
a whole (hours/month)				
110. Application production	x			
111. Application preparation	x			
112. Program development and				
maintenance	x			
113. Total chargeable lost time	x			
114. Total operational use				
(Σ 110 through 114)	x			
115. Total nonchargeable lost				
time	x			
116. Monthly rental, $	x			
For application:				
117. Production hours/month	x			
% of production allocatable				
118. Input edit	x			
119. File maintenance	x			
120. Report generation	x			
121. Merge	x			
122. Sort	x			
123. Compute	x			
124. Query	x			
125. Program development and				
maintenance hrs/month	x			

F. SUMMARY SHEET FOR NUMERICAL DATA
(Continued)

System____________________ Location________________________ Analyst_________

Date____________

Name of Variable	Must Collect	Collect if Applicable	Collect if Easily Obtainable	Value of Variable or Remarks
126. Hours for compilation, assembly, checkout, system test during development phase	x			
Processing functions				
127. Total object instruction in application	x			
128. Total source statements in application		x		
129. Total object instructions in executive		x		
130. Planning information (any personnel cost, development time, hardware cost, or processing function variables stated at time of proposal)[1]			x	

[1] Must be combined with other variables in regression analysis.

APPENDIX C

DEFINITIONS OF FACTORS AND DESCRIPTORS

A. Dependent Variables

Symbol	Name	Definition
Y_1	Man-months of development effort	The number of man-months expended by all relevant personnel including managers, analysts, programmers, and operators to develop the ADPS during the development phase which begins with the start of system design and ends when the system is declared operational. During this development phase, such activities as detailed system design, programming, checkout, and equipment installation are accomplished.
Y_2	Dollars of hardware cost for program checkout	The hardware cost for computer hours used for program checkout during the development phase of the ADPS.
Y_3	Number of program maintenance personnel	The number of personnel, including managers, analysts, and programmers, involved in improving, changing, and correcting programs of a system during the operations phase.
Y_4	Number of operations personnel	The number of related personnel, including operators, schedulers, data edit personnel, magnetic tape librarians, report binders, managers, etc., used to process the ADPS programs on the computer during the operations phase.
Y_5	Dollars per month of hardware cost for application production	The hardware cost for monthly computer hours charged to the user of the ADPS for processing that is not of a developmental or corrective nature.
Y_6	Dollars per month of hardware cost for program maintenance	The hardware cost for monthly computer hours used for processing improvements, changes, and corrections to programs of an operational ADPS.
Y_7	Months of elapsed development time	The number of calendar months elapsed from the date system design for the ADPS is begun to the date it is declared operational.

Symbol	Name	Definition
Y_8	Number of source statements	The number of lines of code written by the programmer in any source language for the ADPS. This may be the same as the number of instructions in machine language.
Y_9	Number of object instructions	The number of instructions generated by the compiler or assembler for the ADPS. This is the number of machine-format instructions in an object program deck that can be processed directly by the computer.
Y_{10}	Percent of production hours for input edit	The percent of production hours per month for input edit where input edit is performed on input data to prepare it for the primary processing; e.g., limit and logic checking, field conversion, and data edit.
Y_{11}	Percent of production hours for file maintenance	The percent of production hours per month for file maintenance where file maintenance is the modification of a file to incorporate corrections, additions, and deletions.
Y_{12}	Percent of production hours for report generation	The percent of production hours per month for report generation where report generation is the transformation of results from primary computations to outputs for the system user.
Y_{13}	Percent of production hours for merge	The percent of production hours per month for merge where merge is the combining of items of records from two or more sequenced files with the same key into one sequenced file.
Y_{14}	Percent of production hours for sort	The percent of production hours per month for sort where sort is the arranging of records of information according to rules operating upon key(s) contained in the records.
Y_{15}	Percent of production hours for compute	The percent of production hours per month for compute where compute is the performance of logical, arithmetic, and decisional operations on data.

Symbol	Name	Definition
Y_{16}	Percent of production hours for query	The percent of production hours per month for query where query is acting on a demand input which specifies that data be accessed via file search and be displayed or output.
Y_{17}	Percent of production hours for control	The percent of production hours per month for control where control is a computer processing function that expedites all other computer processing functions; e.g., job scheduling, priority handling, segment overlaying, data management, and hardware assignment, etc.

B. Independent Variables

Symbol	Name	Definition
X_1	Characters per month of input volume	The expected amount of ADPS input originating outside the ADPS, measured in characters per month. Intermediate inputs of the ADPS should not be included. On unit record input, only character positions used for data are counted.
X_2	Number of input transaction types	A count of different transaction types of ADPS input which normally are identified by a unique transaction code and/or a unique input format.
X_3	Number of input data fields	A count of data fields from the ADPS input that are unique in content and/or format; e.g., if there is a data field for name on six different card formats, the number of unique data fields is one.
X_4	Percent of input rejects	Input data error rate measured by the ratio of the number of rejected records to the number of expected records per month multiplied by 100.
X_5	Characters per month of output volume	The expected amount of ADPS output destined to users, measured in characters per month. Intermediate outputs of the ADPS are not included. Only nonblank characters are counted.
X_6	Number of output formats	The number of different types and formats of ADPS outputs.

Symbol	Name	Definition
X_7	Characters in data base	The expected number of characters in the data base where the data base is a collection of files that contain unique information, are accessible to the ADPS, and are normally referenced or updated with relatively high frequency. Intermediate files are not counted.
X_8	Number of data base record types	The number of logical record types in the data base where a logical record is a set of logically related data fields independent of the physical manner of storage.
X_9	Percent of source statements for input edit	The percent of source statements for input edit. (See Y_8 for definition of source statements and Y_{10} for definition of input edit.)
X_{10}	Percent of source statements for file maintenance	The percent of source statements for file maintenance. (See Y_{11} for definition of file maintenance.)
X_{11}	Percent of source statements for report generation	The percent of source statements for report generation. (See Y_{12} for definition of report generation.)
X_{12}	Percent of source statements for merge	The percent of source statements for merge. (See Y_{13} for definition of merge.)
X_{13}	Percent of source statements for sort	The percent of source statements for sort. (See Y_{14} for definition of sort.)
X_{14}	Percent of source statements for compute	The percent of source statements for compute. (See Y_{15} for definition of compute.)
X_{15}	Percent of source statements for query	The percent of source statements for query. (See Y_{16} for definition of query.)
X_{16}	Percent of source statements for control	The percent of source statements for control. (See Y_{17} for definition of control.)
X_{17}	Average number of years of college education for development managers	Development managers college education, measured in average number of years, where development managers are the individuals responsible for directing and coordinating all or part of the activities associated with an

Symbol	Name	Definition
		ADPS during the development phase. Only managers devoting at least 10 percent of their time to the system are considered.
X_{18}	Average number of years of ADP experience for development managers	Average number of years in the field of automatic data processing (ADP) for development managers. (See X_{17} for definition of development managers.)
X_{19}	Average number of years of ADP experience for analysts	Average number of years in ADP for analysts, who are persons skilled in the definition of and the development of techniques for solving a problem.
X_{20}	Average number of years of ADP experience for programmers	Average number of years in ADP for programmers, who are persons who prepare problem solving procedures and logical flow charts, and code and debug programs.
X_{21}	Average number of years of ADP experience for operations personnel	Average number of years in ADP for operations personnel. (See Y_4 for definition of operations personnel.)
X_{22}	Average number of years of functional area experience for development managers	Average number of years of experience in a field of application, such as accounting, inventory control, weather forecasting, etc., for development managers. (See X_{17} for definition of development managers.)
X_{23}	Average number of years of functional area experience for analysts	Average number of years of experience in a field of application for analysts. (See X_{19} for definition of analysts.)
X_{24}	Average number of years of functional area experience for programmers	Average number of years of experience in a field of application for programmers. (See X_{20} for definition of programmers.)
X_{25}	Average number of years of functional area experience for operations personnel	Average number of years of experience in a field of application for operations personnel. (See Y_4 for definition of operations personnel.)

Symbol	Name	Definition
X_{26}	Months of machine maturity	The number of calendar months between the first delivery date for the model of base machine used by the ADPS and the date of initial checkout of the ADPS computer programs. The delivery date is given by Adams Associates, Computer Characteristics Quarterly; the April 1966 edition was used for the purpose of this study.

APPENDIX D

COLLECTED DATA

(INCLUDING RELIABILITY STATEMENTS, MEANS, AND STANDARD DEVIATIONS)

PHASE II COLLECTED DATA AND STATISTICS

Class	Symbol	Name	Management Supporting Data Systems--Logistics						Management Supporting Data Systems--Personnel/Finance						Operations Supporting Data Systems			Research and Development Supporting Systems			Mean	Standard Deviation	Sample Size
			GE/BSS	DSWC	MILSTAMP	PDS	1050/HSS	RRC	AMPS	SC/ACCT	MAFR	RAFT	PDSO/MPC	PDSO/MAC	TCC	GWC	SPCTRK	ADOBE	ORBIT	MISSIM			
Dependent Variables																							
Cost factor	Y_1	Man-months of development effort	375 (2)	86 (2)	61 (1)	267 (1)	1,400 (2)	118 (3)	704 (3)	226 (2)	392 (3)	381 (2)	1,443 (1)	489 (3)	748 (2)	4,404 (3)	2,424 (3)	21 (3)	49 (3)	66 (3)	758.50	1,107.00	18
	Y_2	Dollars of hardware cost for program checkout	212,000 (4)	140,660 (3)	70,847 (4)	151,780 (2)	Unknown	62,190 (3)	37,178 (3)	28,525 (3)	73,900 (1)	29,970 (1)	Unknown	84,700 (3)	Unknown	Unknown	Unknown	Unknown	18,050 (2)	Unknown	82,710	61,200	11
	Y_3	Number of program maintenance personnel	4 (2)	5 (2)	2 (1)	8 (1)	68 (1)	6 (2)	8 (1)	2 (2)	11 (1)	4 (1)	50 (2)	34 (1)	17 (3)	59 (2)	20 (1)	1.2 (2)	0 (1)	1.2 (3)	16.69	21.40	18
	Y_4	Number of operations personnel	8 (2)	2 (1)	4.5 (3)	10 (3)	7.7 (1)	1 (1)	3.5 (2)	Unknown	8.3 (2)	3 (1)	52 (2)	17 (2)	11 (2)	91 (1)	60 (2)	1 (2)	0.1 (1)	2.1 (1)	16.60	25.60	17
	Y_5	Dollars per month of hardware cost for application production	9,400 (2)	6,671 (2)	16,250 (2)	16,870 (2)	12,598 (2)	1,906 (2)	1,725 (2)	13,448 (2)	17,553 (2)	8,190 (2)	49,640 (2)	8,840 (2)	20,921 (2)	178,600 (2)	181,400 (2)	3,331 (2)	140 (2)	30,233 (2)	32,100	55,080	18
	Y_6	Dollars per month of hardware cost for program maintenance	267 (4)	4,289 (2)	2,030 (2)	484 (1)	Unknown	953 (2)	0 (1)	2,282 (2)	14,706 (2)	585 (4)	15,140 (2)	1,617 (2)	15,133 (2)	61,697 (2)	76,423 (2)	1,032 (2)	0 (1)	1,333 (3)	11,640	22,430	17
Other factors	Y_7	Months of elapsed development time	10 (1)	12 (2)	11 (2)	18 (1)	18 (1)	4 (3)	18 (1)	12 (1)	7 (1)	23 (1)	20 (1)	22 (2)	25 (1)	21 (4)	45 (4)	25 (4)	24 (2)	38 (2)	19.60	10.21	18
	Y_8	Number of source statements	150,000 (3)	208,100 (2)	24,310 (1)	14,200 (1)	Unknown	19,300 (2)	14,140 (1)	39,360 (2)	25,240 (4)	34,620 (1)	195,100 (1)	48,660 (1)	150,000 (2)	99,790 (1)	423,300 (2)	4,532 (2)	4,580 (2)	21,380 (2)	86,860	110,400	17
	Y_9	Number of object instructions	150,000 (3)	208,100 (2)	97,250 (4)	14,200 (1)	340,000 (3)	19,300 (2)	14,140 (1)	66,000 (2)	101,000 (1)	34,620 (1)	780,500 (4)	219,000 (4)	150,000 (2)	124,200 (1)	485,200 (2)	22,500 (2)	16,020 (2)	84,000 (4)	162,500	198,500	18
Application for production factor	Y_{10}	Percent of production hours for input edit	0 (4)	0 (2)	Unknown	Unknown	Unknown	4 (3)	7 (3)	5 (4)	12 (2)	17 (1)	36 (2)	7 (1)	Unknown	23 (1)	1 (3)	30 (3)	3 (4)	40 (4)	13.21	13.76	14
	Y_{11}	Percent of production hours for file maintenance	70 (4)	20 (2)	Unknown	Unknown	Unknown	8 (3)	52 (3)	50 (4)	19 (2)	33 (1)	33 (2)	18 (1)	Unknown	0 (1)	28 (3)	0 (3)	0 (4)	0 (4)	23.64	22.17	14
	Y_{12}	Percent of production hours for report generation	25 (4)	45 (2)	Unknown	Unknown	Unknown	10 (3)	34 (3)	30 (4)	3 (2)	26 (1)	10 (2)	31 (1)	Unknown	31 (1)	6 (3)	17 (3)	20 (4)	10 (4)	21.28	12.39	14
	Y_{13}	Percent of production hours for merge	2 (4)	0 (2)	Unknown	Unknown	Unknown	0 (3)	0 (3)	0 (4)	2 (2)	6 (1)	1 (2)	3 (1)	Unknown	0 (1)	1 (3)	0 (3)	0 (4)	0 (4)	1.07	1.73	14
	Y_{14}	Percent of production hours for sort	3 (4)	35 (2)	Unknown	Unknown	Unknown	67 (3)	7 (3)	5 (4)	44 (2)	3 (1)	4 (2)	22 (1)	Unknown	0 (1)	2 (3)	0 (3)	0 (4)	0 (4)	13.71	20.78	14
	Y_{15}	Percent of production hours for computer	0 (4)	0 (2)	Unknown	Unknown	Unknown	9 (3)	0 (3)	0 (4)	20 (2)	13 (1)	0 (2)	1 (1)	Unknown	45 (1)	56 (3)	53 (3)	75 (4)	50 (4)	23.00	26.80	14
	Y_{16}	Percent of production hours for query	0 (4)	0 (2)	Unknown	Unknown	Unknown	2 (3)	0 (3)	9 (4)	0 (2)	2 (1)	16 (2)	18 (1)	Unknown	0 (1)	5 (3)	0 (3)	0 (4)	0 (4)	3.71	6.21	14
	Y_{17}	Percent of production hours for control	0 (4)	0 (2)	Unknown	Unknown	Unknown	0 (3)	0 (3)	1 (4)	0 (2)	0 (1)	0 (2)	0 (1)	Unknown	1 (1)	1 (3)	0 (3)	2 (4)	0 (4)	0.36	0.63	14
Independent Variables																							
Workload factor	X_1	Characters per month of input volume	4,000,000 (2)	1,836,000 (2)	57,880,000 (3)	4,390,000 (3)	7,400,000 (2)	79,570,000 (2)	940,800 (3)	576,000 (2)	18,880,000 (3)	6,402,000 (2)	46,460,000 (2)	11,400,000 (3)	24,650,000 (3)	715,400,000 (3)	26,560,000 (2)	625,800 (3)	480,000 (4)	77,580,000 (2)	$60,280 \times 10^3$	$165,600 \times 10^3$	18
	X_2	Number of input transaction types	42 (1)	38 (2)	3 (3)	11 (3)	229 (2)	10 (2)	14 (1)	334 (2)	65 (2)	48 (1)	122 (1)	209 (1)	58 (4)	24 (2)	41 (2)	5 (2)	3 (2)	2 (1)	72.05	94.41	18
	X_3	Number of input data fields	465 (2)	237 (2)	60 (1)	53 (1)	1,065 (2)	190 (2)	118 (1)	115 (1)	61 (2)	227 (2)	681 (1)	297 (2)	193 (2)	418 (2)	250 (2)	26 (2)	23 (2)	15 (2)	251.90	268.30	18
	X_4	Percent of input rejects	1.0 (3)	0.1 (1)	8 (3)	12 (3)	1 (4)	0.1 (1)	0.5 (1)	0.1 (2)	1.9 (1)	0.8 (3)	4.3 (1)	5.6 (3)	5 (3)	4.5 (3)	5 (2)	0.4 (3)	5 (4)	1 (4)	3.13	3.30	18
	X_5	Characters per month of output volume	40,310,000 (3)	13,070,000 (2)	11,950,000 (3)	9,590,000 (3)	7,400,000 (4)	10,310,000 (2)	1,007,000 (3)	756,000 (3)	401,600,000 (3)	26,070,000 (2)	81,600,000 (3)	69,210,000 (3)	5,007,000 (3)	598,000,000 (2)	1,240,000,000 (4)	1,094,000 (3)	3,600,000 (4)	47,130,000 (2)	$142,600 \times 10^3$	$316,900 \times 10^3$	18
	X_6	Number of output formats	43 (1)	33 (2)	21 (1)	20 (2)	339 (2)	11 (1)	11 (1)	27 (1)	156 (2)	57 (1)	141 (1)	21 (2)	31 (4)	121 (1)	70 (2)	11 (2)	2 (3)	63 (2)	65.44	82.28	16
	X_7	Characters in data base	27,000,000 (2)	17,270,000 (3)	28,680,000 (4)	116,400,000 (3)	3,981,000 (4)	627,500 (3)	13,210,000 (3)	3,800,000 (3)	28,550,000 (3)	17,510,000 (3)	276,700,000 (2)	11,570,000 (2)	90,290,000 (3)	3,790,000 (2)	44,420,000 (2)	0 (1)	0 (1)	0 (1)	$37,990 \times 10^3$	$67,560 \times 10^3$	18
	X_8	Number of data base record types	11 (1)	3 (1)	3 (2)	8 (2)	115 (2)	2 (1)	3 (1)	6 (1)	11 (1)	8 (2)	6 (1)	5 (1)	33 (2)	7 (2)	6 (1)	0 (1)	0 (1)	0 (1)	12.61	26.61	18
Complexity factor	X_9	Percent of source statements for input edit	0 (4)	0 (2)	16 (2)	12 (2)	4 (3)	9 (2)	13 (2)	3 (3)	9 (2)	6 (1)	8 (2)	16 (1)	17 (4)	9 (2)	8 (2)	40 (3)	17 (2)	66 (2)	14.05	15.78	18
	X_{10}	Percent of source statements for file maintenance	60 (4)	34 (2)	27 (2)	68 (2)	63 (3)	2 (2)	62 (2)	41 (3)	15 (2)	30 (1)	27 (2)	35 (1)	8 (4)	0 (2)	12 (2)	0 (3)	0 (2)	0 (2)	26.89	24.11	18
	X_{11}	Percent of source statements for report generation	20 (4)	66 (2)	54 (2)	16 (2)	18 (3)	54 (2)	22 (2)	42 (3)	28 (2)	26 (1)	25 (2)	38 (1)	16 (4)	42 (2)	14 (2)	15 (3)	26 (2)	13 (2)	29.72	15.96	18
	X_{12}	Percent of source statements for merge	10 (4)	0 (2)	0 (2)	0 (2)	1 (3)	0 (2)	0 (2)	0 (3)	2 (2)	3 (1)	1 (2)	1 (1)	0 (4)	0 (2)	1 (2)	0 (3)	0 (2)	0 (2)	1.05	2.39	18
	X_{13}	Percent of source statements for sort	10 (4)	0 (2)	0 (2)	2 (2)	3 (3)	0 (2)	2 (2)	3 (3)	8 (2)	3 (1)	4 (2)	4 (1)	2 (4)	0 (2)	5 (2)	0 (3)	0 (2)	0 (2)	2.56	2.89	18
	X_{14}	Percent of source statements for compute	0 (4)	0 (2)	3 (2)	0 (2)	7 (3)	27 (2)	0 (2)	0 (3)	38 (2)	27 (1)	0 (2)	1 (1)	35 (4)	44 (2)	51 (2)	45 (3)	53 (2)	21 (2)	19.56	20.46	18
	X_{15}	Percent of source statements for query	0 (4)	0 (2)	0 (2)	0 (2)	4 (3)	8 (2)	0 (2)	6 (3)	0 (2)	5 (1)	35 (2)	5 (1)	4 (4)	0 (2)	3 (2)	0 (3)	0 (2)	0 (2)	3.89	8.20	18
	X_{16}	Percent of source statements for control	0 (4)	0 (2)	0 (2)	2 (2)	0 (3)	0 (2)	1 (2)	5 (3)	0 (2)	0 (1)	0 (2)	0 (1)	18 (4)	5 (2)	6 (2)	0 (3)	4 (2)	0 (2)	2.28	4.46	18
Education and experience factor	X_{17}	Average number of years of college education for development managers	3.5 (2)	Unknown	4 (1)	2 (1)	Unknown	4 (3)	4.5 (1)	0.5 (1)	4 (1)	4 (3)	3.5 (1)	3.5 (2)	3.5 (1)	6.5 (1)	5 (4)	3 (2)	4 (1)	4 (1)	3.72	1.28	16
	X_{18}	Average number of years of ADP experience for development managers	9 (2)	20 (2)	9 (1)	8 (1)	2.5 (2)	20 (3)	12.5 (1)	7 (1)	11 (1)	1 (3)	8.5 (1)	11 (2)	5 (2)	5 (1)	8 (4)	10.5 (2)	8 (1)	15 (1)	9.50	5.10	18
	X_{19}	Average number of years of ADP experience for analysts	5 (2)	11 (2)	9 (1)	5.5 (1)	0.5 (2)	12 (3)	14.5 (1)	17 (1)	4.5 (1)	2 (3)	4 (1)	8 (2)	4.5 (2)	2.5 (1)	5 (4)	None	8 (1)	12 (1)	7.35	4.65	17
	X_{20}	Average number of years of ADP experience for programmers	4 (2)	11 (2)	7 (1)	7 (1)	2 (2)	2 (4)	4 (1)	0.5 (1)	6 (1)	1.5 (3)	7 (1)	9.5 (2)	3.5 (2)	2.5 (1)	3 (4)	3.5 (2)	4 (1)	6 (1)	4.67	2.83	18
	X_{21}	Average number of years of ADP experience for operations personnel	5 (2)	2 (3)	Unknown	Unknown	Unknown	2 (3)	2 (3)	4 (1)	5.5 (3)	6 (3)	5 (1)	6 (2)	6 (2)	1.5 (1)	2 (3)	5 (2)	8 (1)	Unknown	4.28	2.04	14
	X_{22}	Average number of years of functional area experience for development managers	8 (2)	20 (2)	4 (1)	6.5 (1)	15 (2)	20 (3)	17 (1)	1.5 (1)	20.5 (1)	11 (3)	7 (1)	7.5 (2)	2.5 (3)	13.5 (1)	4 (4)	3.5 (2)	8 (1)	15 (1)	10.25	6.42	18
	X_{23}	Average number of years of functional area experience for analysts	12 (2)	11 (2)	4 (1)	5 (1)	4 (2)	12 (3)	21.5 (1)	17 (1)	12.5 (1)	7 (3)	11.5 (1)	14.5 (2)	1.5 (3)	13 (1)	4 (4)	None	8 (1)	11 (1)	9.97	5.29	17
	X_{24}	Average number of years of functional area experience for programmers	3.5 (2)	11 (2)	1.5 (1)	1.5 (1)	4 (2)	2 (4)	1.5 (1)	0.5 (1)	6 (1)	0 (3)	5.5 (1)	4 (2)	1.5 (3)	9.5 (1)	2 (4)	2 (2)	4 (1)	6 (1)	3.67	3.00	18
	X_{25}	Average number of years of functional area experience for operations personnel	3 (2)	2 (3)	Unknown	Unknown	Unknown	2 (3)	2 (3)	4 (1)	Unknown	0 (3)	1 (1)	2 (2)	2.5 (3)	7 (1)	2 (3)	3.5 (2)	8 (1)	Unknown	2.85	2.24	13
Machine maturity factor	X_{26}	Months of machine maturity	15 (2)	28 (2)	18 (2)	39 (2)	7 (2)	42 (2)	27 (2)	20 (2)	38 (2)	29 (2)	5 (2)	52 (2)	48 (2)	6 (2)	15 (2)	12 (2)	48 (2)	32 (2)	26.72	15.33	18

Notes: (1) Data obtained from a document involving little or no judgment or a judgment made by an authority at the site interpreting a document.
(2) Data obtained from a document or group of documents involving some degree of judgment or interpretation.
(3) Data obtained by inference by the data collector or the authority on the site and not backed up directly by a document.
(4) Information not obtained from the authority and totally based on judgment.

APPENDIX E

METHODOLOGY USED IN THE COST REGRESSION ANALYSIS

A. Introduction

This appendix will set forth the computational formulas employed in obtaining the quantitative results of the statistical analysis. The actual computations were run on a Control Data 3600 computer. The computer programs used to perform these computations are part of the BIOMED package programmed at the UCLA Medical Center (see Reference 1) and are designated by the three letters "BMD" followed by two numbers and another letter (i. e., BMD02R).

B. Scatterplots

Scatter diagrams were obtained by using the plot option available on BMD02D, Correlation with Transgeneration. Plots are made of pairs of values (X_{ij}, X_{ik}) where the value of X_{ij} is plotted on the horizontal axis and the value of X_{ik} is plotted on the vertical axis, and where

$j, k = 1, 2, \cdots, p$

$i = 1, 2, \cdots, n$

p = number of variables

n = number of systems

For a single plot, j and k are held constant, and i varies $1, 2, \cdots, n$.

C. Logarithmic Transformation

To perform a logarithmic transformation, X_{ij} is transformed to the equivalent value in the logarithm to the base 10 scale; in other words,

$$X_{ij} \rightarrow \log_{10} X_{ij} = X'_{ij}$$

Most BMD programs have a feature that allows the user to transform the desired variables before the statistical analysis is performed. In addition, BMD09S, Transgeneration, will transform desired variables and give as output a card deck of the transformed values.

D. Correlation

The correlation coefficient, r_{jk} , as well as the individual variable means and standard deviations, was obtained by using BMD03D, Correlation with Item Deletion.

Let X_{ij} be the jth variable of the ith system, where $(i = 1, 2, \cdots, n)$, $(k, j = 1, 2, \cdots p)$, n is the number of systems, and p is the number of variables. For each X_{ij} value that is accepted for inclusion in the computation, the following steps are performed:

1. Means

$$X_{.j} = \frac{1}{n} \sum_i X_{ij}$$

2. Standard Deviations

$$s_{X_j} = \sqrt{\frac{\sum_i (X_{ij} - X_{.j})^2}{n-1}}$$

3. Correlation Coefficients

$$r_{jk} = \frac{\sum_i (X_{ij} - X_{.j})(X_{ik} - X_{.k})}{\sqrt{\sum_i (X_{ij} - X_{.j})^2 \sum_i (X_{ik} - X_{.k})^2}}$$

E. Analysis of Variance

Analysis of variance techniques applied to test for differences among the means of two or more populations were t-tests, one-way analysis of variance, and analysis of covariance (see References 2 and 3).

1. t-Tests

t-Tests are used to test samples from two populations to determine if the means of the two populations, μ_1 and μ_2, are equal. The assumption made is that both populations have normal distributions with the same mean and the same variance. Then the statistic t_c has a $t(N_1 + N_2 - 2)$ distribution. The computation formula is:

$$t_c = \frac{\overline{X}_1 - \overline{X}_2}{S_p\sqrt{\left(\frac{1}{N_1}\right) + \left(\frac{1}{N_2}\right)}}$$

where

$$S_p^2 = \frac{(N_1 - 1)S_1^2 + (N_2 - 1)S_2^2}{N_1 + N_2 - 2}$$

2. <u>One-Way Analysis of Variance</u>

One-way analysis of variance comprises tests for differences among the means of two or more populations. In general, the hypothesis is $\mu_1 = \mu_2 = \cdots, = \mu_k$ (the means of all categories are equal). The assumption is that the observations are randomly selected from normal populations with homogeneous variance. Then the statistic F_c has a distribution of $F(k-1, N-k)$. The computation formula is:

$$F_c = \frac{\sum_{i=1}^{k} \dfrac{N_i(\overline{X}_i - \overline{X})^2}{k-1}}{\dfrac{\sum_{i=1}^{k}\sum_{j=1}^{N_i} X_{ij}^2 - \sum_{i=1}^{k}\left(\sum_{j=1}^{N_i} X_{ij}^2 \Big/ N_i\right)}{N-k}}$$

The computations were made using BMD01V, Analysis of Variance for One-Way Design.

3. <u>Analysis of Covariance</u>

An analysis of covariance is performed by computer program BMD04V, Analysis of Covariance with Multiple Covariates. This program is designed to compute analysis of covariance information for k subpopulation of Y values, where Y depends linearly on a set of simultaneously observed variables, $X_1, X_2, \cdots, X_p$. The hypothesis being tested in analysis of covariance is stated as follows: There is no difference in the means of the Y values among groups after the Y values have been adjusted according to the X values.

The analysis of covariance test assumes that the regression curves in the k populations are parallel straight lines and the population variances about the regression lines are equal in each of the k populations.

F. Multiple Regression

1. Estimation Equations

The general linear model and the requirements for estimation efficiency were discussed in subsections II.B.3 and 4. The general estimation equation is:

$$\hat{Y}_j = a_j + \sum_{1}^{m} b_{ij} X_i$$

where $\hat{Y}_j$ will be the predicted values for the five cost variables and X_i are the five transformed workload descriptors in logarithms (base 10).

The stepwise regression procedure BMD02R (see Reference 4) was used because the procedure provides a judgment on the contribution made by each variable as though it had been the most recent variable entered. Variables incorporated earlier are reexamined at every stage of the regression. A summary of the procedure is given in the following paragraphs.

- Start with the set of causally related independent variables and enter into regression the X_i variable most highly correlated with the cost variable Y_j.
- Regress Y_j on X_i and obtain least-squares equation. Apply F-test for significance.
- Calculate partial correlation coefficients of all variables not in regression with cost. Choose as the next variable to enter into the regression the one with the highest partial correlation coefficient X_k.
- Develop regression equation $\hat{Y} = f(X_i, X_k)$ by least squares. Apply F-test. Examine contribution of X_i if X_k had been entered first. Apply F-test to X_k and retain X_k if significant.
- Repeat step 3 and choose next variable X_e. Develop regression $\hat{Y} = f(X_i, X_k, X_e)$ by least squares.

Partial F-tests are applied to X_i, X_k and, if significant, they are retained in the regression equation. If additional variables remain, this procedure is continued until no more variables will be admitted to the equation and no more are rejected.

After each variable is added, the following test is performed until the following hypothesis is rejected: The X_{i+1} term makes a significant addition to the regression equation.

$$F_c = \frac{\begin{pmatrix}\text{Regression sum of squares for}\\ \hat{Y} = f(X_1, X_2, \cdots X_i, X_{i+1})\end{pmatrix} - \begin{pmatrix}\text{Regression sum of squares for}\\ \hat{Y} = f(X_1, X_2, \cdots X_i)\end{pmatrix}}{\begin{pmatrix}\text{Residual mean square for}\\ \hat{Y} = f(X_1, X_2, \cdots, X_i)\end{pmatrix}}$$

where F_c is an F-distribution with (1, N-p-1) degrees of freedom; N is the number of observations in the sample and p is number of workload descriptors in the equation.

The result is a vector of parameters $a_j, b1, b2, \cdots, b_m$ that provides the best set of estimators for $\alpha_j, \beta1, \beta2, \cdots, \beta_m$ of the linear regression model. In practice, it is desirable to keep the number of X_is in the estimating relationship with Y_j small because the user prefers a minimum of effort in estimating cost.

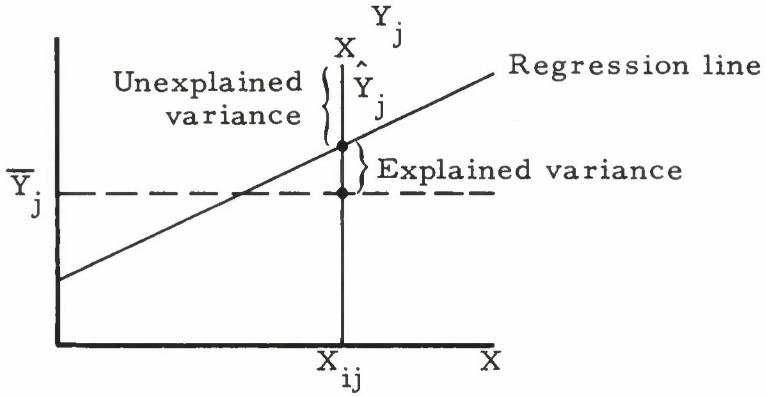
The computer program BMD02R, Stepwise Regression, computes and outputs the following statistics at each step:

- o Multiple correlation coefficient, R
- o Standard error of estimate of $\hat{Y}_j$, $\overline{s}_j$
- o Analysis of variance table
- o For variables in the equation:
  - a. Regression coefficient
  - b. Standard error of regression coefficient
- o For variables not in the equation, partial correlation coefficient

2. <u>Reliability</u>

Total variance pertains to the deviations of the sample Y's from their mean. Explained variance refers to the deviations from $\overline{Y}_j$ of the computed $\hat{Y}_j$ values (calculated from the regression equation) corresponding to the values of X_i in the sample. Unexplained variance is derived from the deviations of the sample Y_j values from the computed values of $\hat{Y}_j$.

$$\sigma_{tj} = \sigma_{ej} + \sigma_{uj}$$



a. <u>Coefficient of Variation</u>

This is a relative measure for standard error of estimate, which is the square root of the unexplained variance adjusted by the number of degrees of freedom (see Reference 5).

$$C = \frac{\sigma_{uj}}{\overline{Y}_j} = \frac{\sqrt{\dfrac{\Sigma(\hat{Y}_j - Y_j)^2}{\text{degrees of freedom}}}}{\overline{Y}_j}$$

Typically, $0 \le C \le 0.2$ is desirable.

b. <u>Coefficient of Correlation</u>

The correlation coefficient is a measure of the degree of association between the dependent variable and the explanatory variables. R is defined as the square root of the proportion of total variance that is represented by the explained variance.

$$R = \sqrt{\frac{\sigma_{ej}^2}{\sigma_{tj}^2}} = \sqrt{\frac{\Sigma(\hat{Y}_j - \overline{Y}_j)^2}{\Sigma(Y_j - \overline{Y}_j)^2}}$$

Typically, $1.0 \ge R \ge 0.8$ is desirable.

c. Residual Analysis

Analysis of residuals, the unexplained deviations $\overline{Y}_j - \hat{Y}_j$, is a useful tool for further analysis of the unexplained variance. In addition, it provides analytic tests for truths of assumptions made in regression analysis. These are that errors or deviations are independent, have zero mean and a constant variance σ^2, and are normally distributed. The usual procedure is to plot the residuals as shown in Figure 15 (see Reference 6).

3. Prediction Intervals

The preceding measures are measures of reliability considered in the context of the regression equation in relation to the sample observations. As a measure of predictive efficiency, the concept of the prediction interval is used. For given values of the explanatory variables X_{ij}'s, the estimating equation is used to obtain a predicted value $\hat{Y}_j$. A boundary is placed around $\hat{Y}_j$, $\hat{Y}_j \pm A$, such that there is a certain level of confidence that the established interval brackets the population value of Y_j. An 80 percent prediction interval does not mean that the probability is 0.80 that the population value of Y_j lies between that interval. Rather it means that there is 80 percent confidence, in a subjective sense, that this is the case. This is fiducial probability and not a true probability statement.

In the case of one explanatory variable, this interval typically can be depicted as the area between two hyperbolae, one on each side of the regression line. In the case of two explanatory variables, this interval can be depicted as the space between two hyperboloids, one on each side of the regression plane. For larger numbers of explanatory variables, this interval cannot be graphically portrayed, and the equation for the interval becomes increasingly more complicated (see Reference 7).

The 1-α prediction limits for a Y_j obtained from a particular set of X_i values are given by

$$\hat{Y}_j \pm t_{[N-p-1,\ 1-\frac{1}{2}\alpha]}\ \overline{s}_j \left[1 + \frac{1}{N} + \sum_{i=1}^{p} (X_i - \overline{X}_i)^2 c_{ii} \right.$$

$$\left. + 2 \sum_{i=1}^{N-1} \sum_{k=i+1}^{p} (X_i - \overline{X}_i)(X_k - \overline{X}_k)\ c_{ik} \right]^{1/2}$$

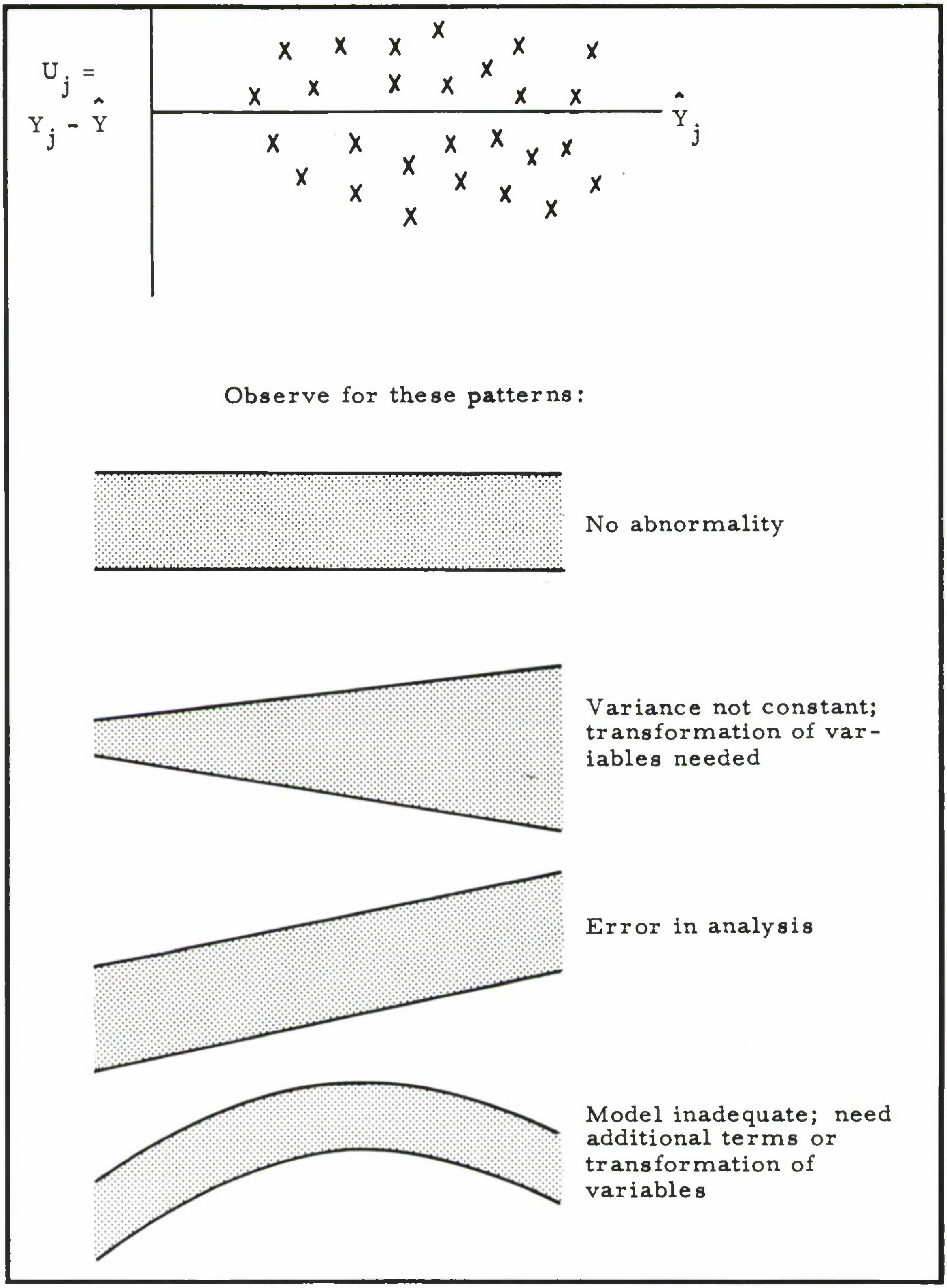


FIGURE 15 - SAMPLE PATTERNS OF RESIDUAL PLOTS

where

$$\begin{array}{lcccccc} & & \underline{i=1} & \underline{i=2} & \cdots & \underline{i=p} \\ c_{i1}\sum v_1^2 + c_{i2}\sum v_1 v_2 + \cdots + c_{ip}\sum v_1 v_p & = & 1 & 0 & \cdots & 0 \\ c_{i1}\sum v_1 v_2 + c_{i2}\sum v_2^2 + \cdots + c_{ip}\sum v_2 v_p & = & 0 & 1 & \cdots & 0 \\ c_{i1}\sum v_1 v_p + c_{i2}\sum v_2 v_p + \cdots + c_{ip}\sum v_p^2 & = & 0 & 0 & \cdots & 1 \end{array}$$

and

$$v_\ell = X_{\ell i} - \overline{X}_\ell \quad \text{for } i = 1, \cdots N$$

$$c_{ik} = c_{ki}$$

$\overline{s}_j$ = adjusted standard error of estimate

G. Factor Analysis

The factor analysis was performed by using BMD03M, General Factor Analysis. This program uses a principle component method with an orthogonal rotation of the factor matrix. Communalities are estimated from the squared multiple correlation coefficients. The complete computational procedure is given in BMD03M. For further details on factor analysis see Reference 8.

H. Hardware Costs

The hardware costs for the 18 systems surveyed were calculated by using the following computation procedure:

1. For program checkout

Let A = total checkout hours for subject ADPS, base machine

S = total checkout hours for subject ADPS, satellite machine

R_B = basic hourly rental, base machine(1)

R_S = basic hourly rental, satellite machine(1)

Y_2 = dollars of hardware cost for program checkout

[1] For purchased components, the applicable GSA monthly rental costs were used.

Then

$$Y_2 = AR_B + SR_S$$

2. For application production or application program maintenance

Let A = monthly hours for ADPS, base machine

B = total monthly hours, base machine

S = total monthly hours, satellite machine

R_B = basic hourly rental, base machine[1]

R_S = basic hourly rental, satellite machine[1]

C_B = central processing unit extra time hourly rate, base machine

C_S = central processing unit extra time hourly rate, satellite machine

Where

Y_5 = dollars per month of hardware cost for application production

Y_6 = dollars per month of hardware cost for program maintenance

<u>Case I</u>: $A < 200$, $AS/B < 200$; then

$$Y_5, Y_6 = \left[AR_B\right] + \left[\frac{AS}{B} R_S\right]$$

<u>Case II</u>: $A < 200$, $AS/B > 200$; then

$$Y_5, Y_6 = \left[AR_B\right] + \left[200 R_S \left(\frac{AS}{B} - 200\right) C_S\right]$$

<u>Case III</u>: $A > 200$, $AS/B < 200$; then

$$Y_5, Y_6 = \left[200 R_B + (A - 200) C_B\right] + \left[\frac{AS}{B} R_S\right]$$

---

[1] For purchased components, the applicable GSA monthly rental costs were used.

Case IV: $A > 200$, $AS/B > 200$; then

$$Y_5Y_6 = \left[200\ R_B + (A - 200)\ C_B\right] + \left[200\ R_S + \left(\frac{AS}{B} - 200\right)\ C_S\right]$$

Note that the preceding procedure is applicable for an installation with at most one base and one satellite computer. If there are two base or satellite computers and:

Case V: $A < 400$, $AS/B < 400$; then use the appropriate Cases I through IV without making any changes.

Case VI: Either $A > 400$ and two base computers are installed or $AS/B > 400$ and two satellite computers are installed, or both; then use the appropriate Cases I through IV, and, where there are two computers, change the constants 200 to 400 as appropriate.

LIST OF REFERENCES

1. Dixon, W.J., BMD Biomedical Computer Programs, Health Sciences Computing Facility, UCLA, September 1965

2. Dixon, W.J. and F.J. Massey, Jr., Introduction to Statistical Analysis, McGraw-Hill Book Company, Inc., 1957

3. Bennett, C.A. and N.L. Franklin, Statistical Analysis in Chemistry and the Chemical Industry, John Wiley and Sons, Inc., 1961

4. Draper, N.R. and H. Smith, Applied Regression Analysis, John Wiley and Sons, Inc., 1966

5. Fisher, G.H., "Use of Statistical Regression Analysis in Deriving Estimating Relationships," Chapter V in Concept and Procedures of Cost Analysis, J.P. Large, Editor, RAND Corporation, RM-3589-PR, June 1963

6. Guildford, J.P., Fundamental Statistics in Psychology and Education, McGraw-Hill Book Company, Inc., 1956

7. Duncan, A.J., Quality Control and Industrial Statistics, 1952

8. Fruchter, Benjamin, Introduction to Factor Analysis, Van Nostrand Company, Inc., 1954

9. Farrar, D. E. and R. E. Apple, Some Factors That Affect the Overhaul Cost of Ships: An Exercise on Statistical Cost Analysis, Naval Research Logistics Quarterly, Volume 10, No. 4, December 1963

APPENDIX F

CURRENT ADPS PROPOSAL PROCEDURES

A. Introduction

One of the major objectives of this contract is to propose tools to the decision makers at HQ USAF to assist them in judging proposals for new automation. For any tool to be constructed in the most useful manner, it is necessary to understand who the decision makers are, what analytical procedures they follow in judging proposals for new automation, and what the form and content of such proposals are. To the extent possible within contract scope, the PRC project team has gathered such data through a study of applicable Air Force regulations and through many lengthy discussions with personnel at HQ USAF.

This appendix summarizes the various regulatory procedures that govern the preparation and submission of proposals involving ADP systems to HQ USAF. It is not claimed that these represent all applicable procedures, but PRC is certain that the majority of all ADPS proposals are covered by the regulations discussed herein. It should be clear, after perusal of this appendix, just how complex the proposal-judging function is and how urgently the decision makers need additional tools.

Specifically, the remainder of this appendix discusses 300 series regulations and the functions of AFADA, 375 and 57 series regulations and system management procedures, 100 series regulations governing communications systems, and AFR 80-2 concerning research and development.

Various organizations within the Air Force are referenced herein and the organization chart presented in Figure 16 should help identify the position of a given organization within the Air Force structure.

B. AFR 300 Series Regulations

This series deals in general with the design, implementation, and operation of automated data systems for management supporting data systems, operations supporting systems, and research and development supporting data systems. It also pertains to the selection, acquisition, and management of automatic data processing equipment for these systems, with the following notable exceptions:

- o Data systems and/or equipment integral to a weapon system
- o ADPS under development for a particular use through the expenditure of research and development test and evaluation funds
- o Analog computing systems

AFR 300-2 establishes the Air Force general objectives and policies in the area of data automation and specifies that the Senior ADP Policy Official for the Air Force is the Assistant Secretary of the Air Force (Financial Management). In this capacity, he is responsible for the

administration of the Air Force ADP program and the selection and acquisition of ADP equipment; accordingly, all proposals for ADP equipment acquisition must be approved by him. AFADA has been designated by SAFFM as the focal point for coordinating and integrating the Air Force data automation effort. Functions performed by AFADA will be covered in subsequent paragraphs.

1. AFR 300-3, Management Supporting Data Systems

This regulation establishes procedures and responsibilities for the design, implementation, modification, and maintenance of management supporting data systems. In most cases a Data Automation Proposal (DAP) is mandatory. Procedures and formats for DAP preparation and submission are included in this regulation. Program control of design and implementation of management supporting data systems is exercised through the Data System Automation Program (DSAP). HQ USAF makes DSAP entries, reflecting the separate design and implementation phases of automated data systems, as follows:

- Systems Development Projects Inventory. This entry reflects issuance of a Data Project Directive and indicates data system design activity by location and scheduled completion date.

- Data System Implementation Schedule. This entry reflects current implementation plans and identification of the support ADP equipment scheduled for each location.

- Current System Inventory. This entry reflects current active data systems and ADP equipment in use in support of such data systems.

Reporting procedures are those outlined in AFM 171-9.

Systems proposed under this regulation are categorized as either standard or unique. Standard data systems are common to two or more commands or agencies and possess uniformity of inputs, file content, processing logic, and outputs. Unique data systems are peculiar to a single command or agency.

HQ USAF (AFADAC) must review DAP's received to determine the following:

- Acceptance, and (a) establishment of a system development project, (b) other directed action prior to implementation, or (c) directed implementation

- Nonacceptance, and (a) return for additional information or development, or (b) return with explanation of nonacceptability

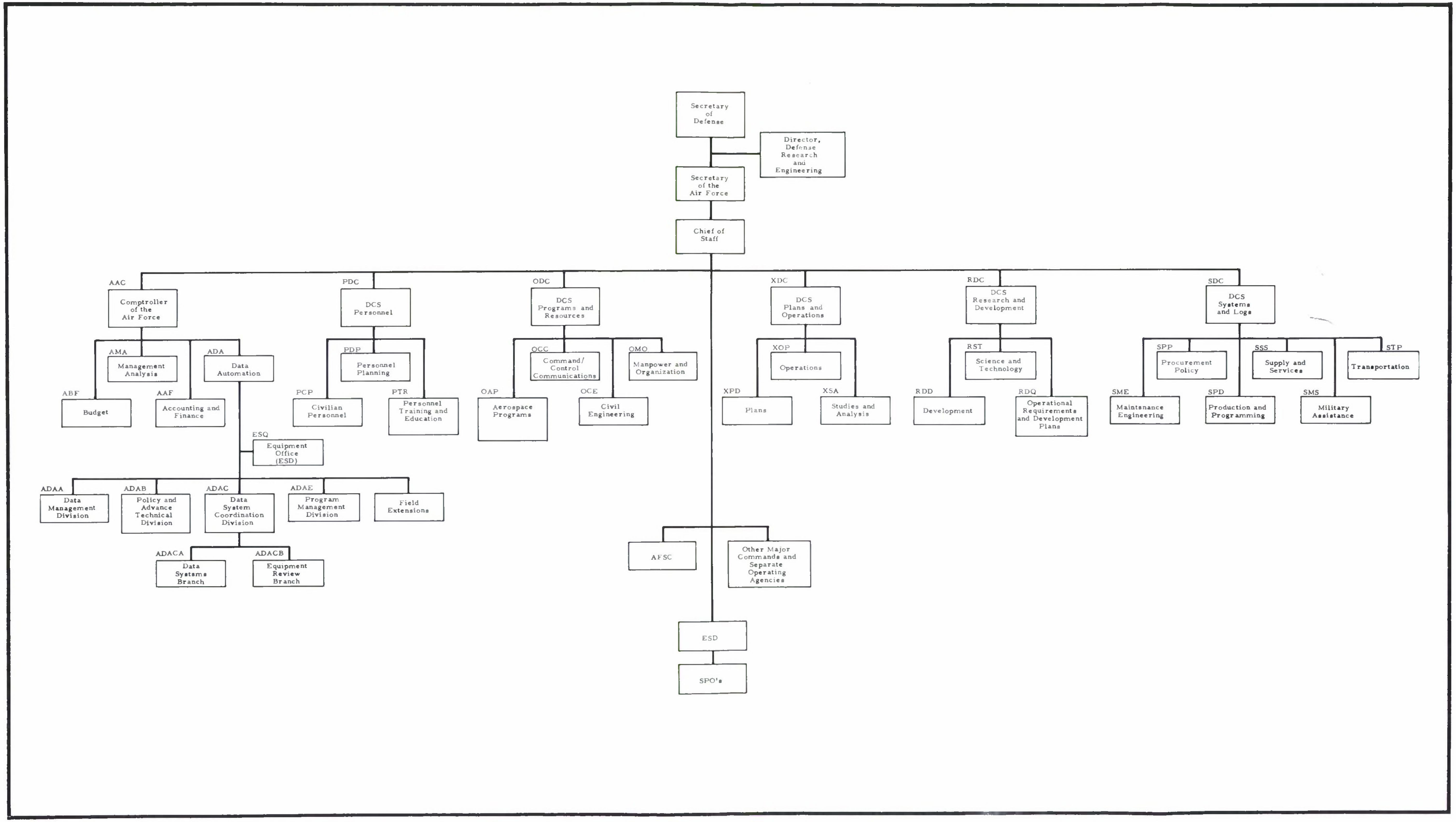


FIGURE 16 - AIR FORCE ORGANIZATION CHART (PARTIAL)

Because AFADA is the decision authority for management, operations, and research and development supporting data systems, something should be said at this point concerning its organization, functions, and overall responsibilities. All of these are covered in detail in AFM 170-6; however, it should prove instructive to describe those functions associated with the approval process for DAP's.

Figure 17 shows the organization of AFADA. All DAP's go to AFADACA for coordination and evaluation. It is their responsibility to see that all interested members of the Air Staff are involved in the evaluation process. Each DAP is logged in and given a number. The goal at AFADACA is to completely process a DAP in no longer than 45 days. The DAP is subjected simultaneously to an in-house review and a functional review. The functional review consists of sending the DAP to any part of the Air Staff which might be involved or interested (e.g., DCS/Personnel if additional manpower is required).

The in-house review consists of sending the DAP to those parts of AFADA which might have some comment, and almost always includes AFADAA, AFADAB, AFADAE, and AFADO. Typical responsibilities of these organizations are as follows:

1. AFADAA. Key, but not all inclusive, responsibilities as described in AFM 170-6 are:

> "Reviews, validates, and has approval authority for all data system content and standard output therefrom (AFR 300 series). Insures standardization of this data to provide interface capabilities and to preclude non-essential overlap or duplication within and between systems and reports.
>
> "Prescribes the system and procedures for a continuous Air Force-wide review, analysis and validation of all reports, data bank content, and standard outputs. Conducts periodic reviews of all reporting requirements placed on the Air Force by other Federal agencies and the public.
>
> "Directs and is responsible for the Air Force Data Elements and Codes Standardization program including the approval, publication and implementation of standard data elements, data items, data codes, data descriptors, and data field designators. Provides guidance and advice to Data Automation Working Groups on these matters. Resolves functional area conflicts.
>
> "Establishes and controls automated file(s) for data elements and related features (data items, codes, descriptors, and field designators), including a repository of the data content of standard data banks and Headquarters USAF directed or implemented reports.

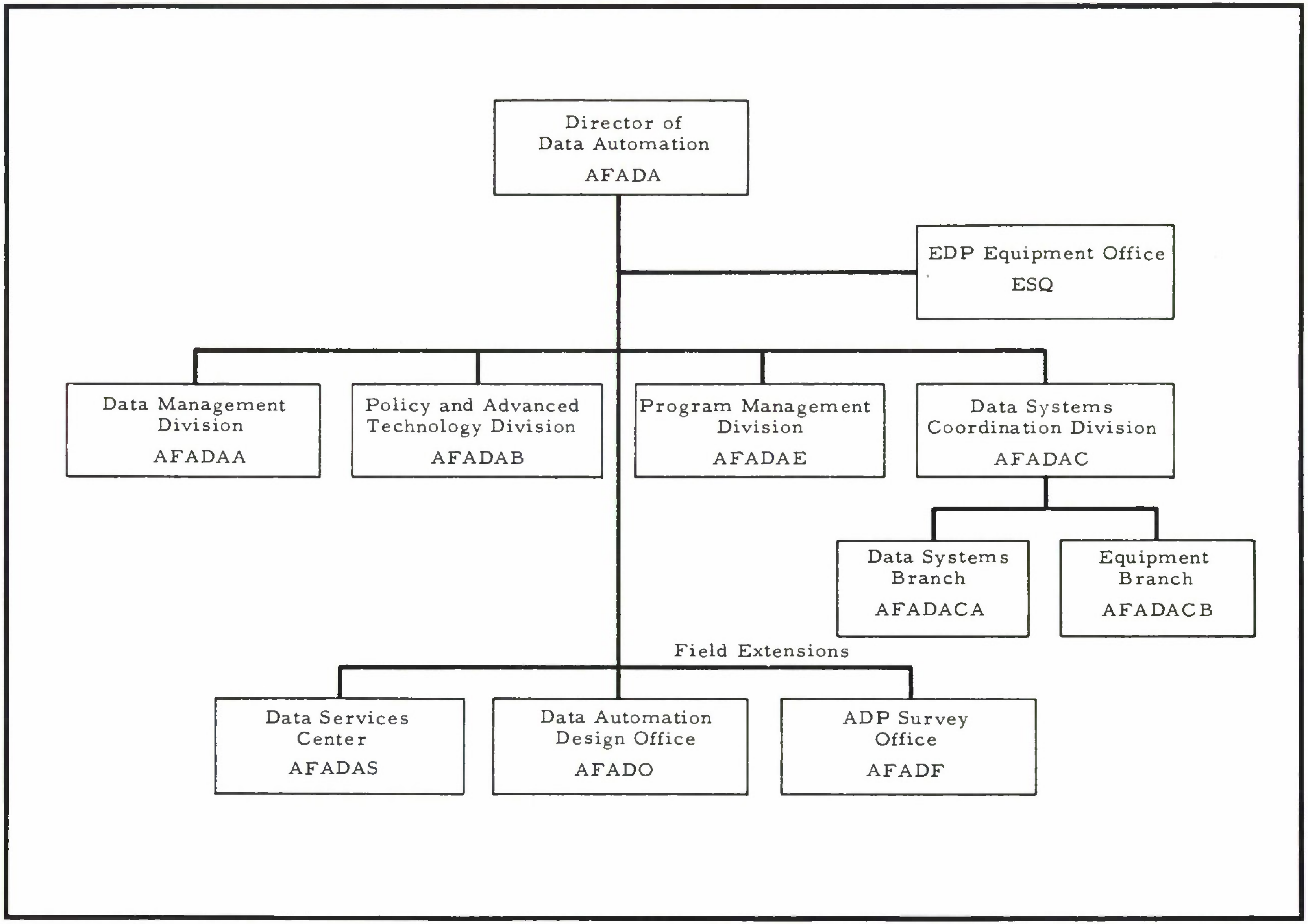


FIGURE 17 - ORGANIZATION OF AFADA

"Evaluates information requirements of the Secretary of the Air Force, Chief of Staff, and other principal Air Staff officers. Assures that valid requirements are in data banks or reports."

Accordingly, AFADAA's main function with respect to DAP review is to insure that reports, data elements, codes, etc., are in compliance with AFR 174-1 and AFR 300-4 as required.

2. AFADAB. Again quoting from AFM 170-6, key responsibilities of this organization include:

"Serves as focal point and is responsible for data automation objectives, concepts, plans and policies in support of overall Air Force objectives and plans.

"Develops the regulatory structure for effective management of the total data automation effort.

"Serves as the Air Force focal point with DOD on all matters pertaining to data automation objectives, concepts and policies, and as the AFADA coordinating office on all DOD matters.

"Establishes and coordinates Air Force requirements for technical data automation studies and development projects; monitors their progress and evaluates results.

"Establishes policies pertaining to data automation technical standards for Air Force use, and coordinates the development and adoption of technical standards with other agencies or industry.

"Plans for the interface and integration of Air Force management and operational supporting data systems to insure efficiency and elimination of duplication."

In reviewing a DAP, AFADAB determines whether regulations in addition to the AFR 300 series should apply and whether established standards are involved or suggested.

3. AFADAE. Key functions as stated in AFM 170-6 include:

"Exercises surveillance over USAF data automation installations; evaluates progress and performance against programs and standards; and initiates corrective action when necessary.

"Plans for and monitors the installation, operation, and management of all ADP Equipment after the equipment selection and approval process has been completed.

"Prescribes and manages the USAF Data Systems Automation Program (DSAP) and changes thereto.

"Reviews requests for ADPE and recommends approval action based on budget requirements and current management actions.

"Reviews and approves requests for ADP services through service contracts.

"Compiles Data Automation program cost, ADPE utilization and inventory data for the Air Staff, OSD, BOB and other Government agencies use.

"Performs continuous post installation studies of method of acquisition of ADPE and initiates purchase action when economically advantageous.

"Administers the relocation or disposition of surplus Government-owned ADP Equipment."

Manpower implications in the DAP are analyzed and discussed.

4. AFADO. This organization determines whether the system proposed in the DAP is unique or standard. It might also recommend holding up a proposed unique system because of some standard system already under development. If a proposed unique system has Air Force-wide benefits, AFADO might establish it as a standard system. AFADO maintains the Air Force's standard Management Supporting Data Systems and normally implements such systems.

The instructions for preparing a DAP are included as Attachment 2 of AFR 300-3. A copy of this attachment is presented in Figure 18. The current instructions call for only additional resources required. Current practice at AFADAC is to request all resources required before a DAP can be properly evaluated.

Several key questions must be answered when evaluating a DAP, all of which are answered, with varying degrees of success, by AFADAC proposal evaluators:

- Does the Air Force need it? In other words, does the proposed ADPS fall within the policies and objectives of the Air Force as a whole and the specific mission of the requestor? This is by far the hardest question to answer and, once answered, the one most subject to argument.

- If a valid mission requirement exists, is the proposed ADPS the best technical and most economical solution? And, as a

DATA AUTOMATION PROPOSAL (DAP) SUBMISSION

General Instructions. Complete detail pertaining to each DAP item may not be available (or required) at the time of DAP submission. However, each item should be completed to the degree appropriate at the time of submission. Items not directly pertinent to the specific proposal should be marked "Not Applicable." The following format must be followed:

1. **Identification.** Indicate originating base and/or organization, parent command, and preparation date.

2. **Title and Purpose.** Identify the data automation requirement/recommendation; specify what is to be accomplished; and relate this to an established function or responsibility; specify the data automation characteristics involved; and indicate any associated organizational and procedural changes contemplated.

3. **System/Modification Description.** Specify the inputs and file content, and provide a general flow diagram showing processing operation. Identify outputs and their relationship with other data systems. Indicate processing workload, responsiveness criteria, etc., at appropriate points within the processing operation.

4. **Resource Requirements.** Indicate, to the degree possible, the anticipated additional resources required (over those now in use) for the proposed system or modification under normal operating conditions. Resource requirements should be specified as being command or Air Force-wide, separately identified within the following groups:
   a. Personnel (grade/man months or years).
   b. Equipment (identify, and include approximate dollar cost).
   c. Physical facilities (site preparation, approximate dollar cost).
   d. Communications (identify number of units, approximate dollar cost).
   e. Other (as appropriate).

5. **Summary of Benefits.** Indicate, to the degree practicable, the economies and/or other benefits to accrue on a command or Air Force-wide basis through the proposed system or modification. Tangible benefits (personnel, equipment, or other savings) should be summarized to indicate an estimated dollar value for a specific time period. Intangible benefits (increased efficiency or responsiveness, accomplishment of tasks not previously feasible or possible, preclusion of increased cost of current operations, etc.) should be outlined in narrative form, with explanation of derivation of the benefit.

6. **Remarks.** Include additional information which would facilitate understanding and evaluation of the submitted DAP. For new Unique Data Systems include a schedule of proposed locations, if applicable.

FIGURE 18 - PRESCRIBED FORMAT FOR DATA AUTOMATION PROPOSALS

corollary to this question, is there an existing Air Force ADPS that will do the job, or do other ADPS proposals in process support or conflict with the subject proposal?

It is in answering these questions that better tools would be most useful to the proposal evaluators. Although they are currently doing an adequate job in this area, they are not equipped to contend with increases in the proposal load and continuing expansion of data processing in the Air Force; current procedures will become increasingly prone to error, and the time to process a proposal will become longer and longer. More than 700 DAP's have been processed by HQ USAF in the last 5 years; of these, over half were submitted within the last 12 months. If the load continues to increase at this rate, better tools and procedures are mandatory.

At present, the tools available to proposal evaluators are essentially a listing of past and current DAP's in numerical order and the Data System Automation Program (DSAP). The officers within AFADAC who perform proposal evaluations have functional areas of responsibility, which minimizes the amount of information with which they must become familiar and remember. However, these procedures can accommodate an increased workload only by adding more people and establishing a finer functional stratification. Furthermore, there are at present no tools, except the experience of the individual officers performing the evaluation, for assessing cost estimates.

Other responsibilities of AFADA covered by this regulation deal with procedures to be followed after a DAP is approved.

In many cases it is deemed desirable to establish a system development project for the design (or modification) of automated data systems, development of associated data system specifications, and demonstration of the operational feasibility of new concepts and techniques. In this event, a Data Project Directive (DPD) is issued by AFADA which provides the charter for command or agency initiation of a system development project. One of the key documents produced by the system development project is the Data System Specifications, which provide a complete description of the specific system, including identification of related standard data systems, pertinent standard data elements and codes, input and output definitions, file and record content, and logical flow diagrams of the functions performed. If the Data System Specifications are approved by HQ USAF, an implementation schedule is prepared and sent to the command or agency, which in turn prepares the following:

- Available ADP equipment capability
- Funding requirements
- Workload confirmation

- Site preparation requirements
- Training requirements
- Verification of benefits

When all approvals have been made, a final implementation plan is developed to ensure orderly and effective implementation of the data system.

2. Operations Supporting Data Systems

ADP systems for operations supporting data systems currently are acquired through AFR 300-3 (DAP's) or AFR 375-1 (ROC's). A draft version of AFR 300-6, which covers this area, is being studied by AFADA; if adopted, these systems will receive uniform treatment.

3. AFR 300-7, Research and Development Supporting Systems

This regulation distinguishes between research and development support and management or operational supporting data systems. It prescribes responsibilities for establishing and providing scientific/computational ADP equipment support required in conjunction with approved research and development activity. Requirements for new or additional ADP equipment needed primarily to support administration and management of research and development programs must be initiated and developed in accordance with AFR 300-3.

Requests are submitted to AFADAC in the form of a letter of transmittal. If new equipment is required, an equipment specification must be attached to the letter of transmittal. The letter must include the following:

- A statement explaining why augmentation of existing ADP equipment cannot satisfy the requirement
- An analysis of the feasibility of sharing equipment with other Air Force or Government agencies
- Justification for special equipment features, etc.
- A description of the tasks and their associated workload (machine hours and additional manpower)

Although format requirements are different from a DAP, the information required is similar. AFADA actions are also similar. They include the following:

- Review and evaluate the requests

- Screen requirements for possible reutilization of available excess Government-owned or -leased ADP equipment

- Forward equipment specifications to ESD, AFSC, for initiation of ADP equipment selection process

- Obtain higher authority approval for waiver of competitive ADP equipment selection, when required

- Advise the major air command to initiate appropriate ADP equipment acquisition action

4. HOI 300-3, Management Supporting Data Systems

This supplements AFR 300-3 and establishes Air Staff responsibilities in accord with DOD Directives 4105.55 and 5100.40. Key functions of AFADA outlined in this document are as follows:

- Develop and maintain a data system designator (short title) system for data system identification

- Ensure standardization and avoid non-essential overlap and duplication of data systems

- Prescribe standard machine programming language(s) to be used

- Maintain and publish the USAF DSAP

- Disseminate periodically status of DAP's, DPD's, and related actions

- Maintain and prepare AFM 300-4, all approved standard data elements and codes

C. AFR 375 and 57 Series Regulations

System management in the Air Force is defined as the process of planning, organizing, coordinating, evaluating, controlling, and directing the combined effort of Air Force contractors and participating organizations to accomplish system program objectives. The documents of primary interest are AFR 375-1 and HOI 375-1, Management of System Programs.

Programs that come under this type of management are defined as follows:

1. Mandatory. All new (or major modifications of existing) production systems, or new engineering and operational systems developments shall be managed according to AFR 375-1 and HOI 375-1 if they fulfill one or both of the following stipulations:

a. The program is rated in the BRICK-BAT category (AFR 70-24).

b. The program is estimated to require total cumulative RDT&E financing in excess of $25 million; or estimated to require a total production investment in excess of $100 million.

2. Otherwise Designated. Other system programs may be designated for this type of management when they possess one or more of the following characteristics:

a. The program significantly affects U.S. military posture.

b. The program is closely related and, when taken collectively, would qualify under dollar thresholds given above.

c. Significant technical problems are anticipated.

d. Unusual organizational complexity or technological advancement is involved.

e. Extensive interdepartmental, national, or international coordination or support is required.

f. Technological risks are involved that may cause difficulties in many functional areas.

g. Unusual difficulties are presented that require expeditious handling to satisfy an urgent requirement.

In general, the purpose of applying systems management is to ensure that efforts by functional activities of the Air Force are accomplished consistent with the objectives of each system program. Complexity, long lead time, extensive resource requirements, and urgent necessity to attain and maintain maximum operational capability are factors that make it mandatory to apply system management procedures.

Until recently, a system project of the type discussed started when a QOR (Qualitative Operational Requirement), SOR (Specific Operational Requirement), OSR (Operational Support Requirement), or ADO (Advanced Development Objective) was written. AFR 57-1, 17 June 1966, establishes the ROC (Required Operational Capability) as the replacement for QOR's, and the RAD (Requirements Action Directive) as the replacement for SOR's, OSR's, and ADO's.

The ROC is a command's official request to HQ USAF for a new or improved operational capability and, although any organizational level may originate such a document, it must be signed by a general officer or a colonel occupying a key staff position.

The RAD is prepared by HQ USAF, signed by a general officer at directorate level; it directs and guides the Air Force actions necessary to translate a required operational capability into an approved and funded program. The RAD is a guidance document, not a funding instrument; however, it transmits the funding information available at the time it is issued.

The focal point within HQ USAF for the coordination of ROC processing is AFRDQ. Key functions performed include the following:

- o Evaluate the requirement and initiate actions to include, but not be limited to, such items as:

  a. Preparing a plan of action to evaluate the need and satisfy or to disapprove the requirement

  b. Initiating and conducting further studies involving system analysis, tradeoffs, cost effectiveness, etc.

  c. Directing and guiding actions required of AFSC, AFLC, and other major air commands through the RAD

- o Evaluate proposed technical approaches submitted by AFSC, AFLC, industry sources, and other commands.

- o Determine the best acceptable approach, with participation of others as necessary, and submit a proposal to appropriate levels of approving authority. An RAD is normally issued within 60 days of receipt of an ROC.

- o Resolve requirements with allied nations and achieve interservice coordination as required.

Once a system project is established under AFR 375-1, AFSPDO becomes the office of primary responsibility (OPR) for establishing policy and coordinating activities within the Air Staff pertaining to system program documentation and its application to system programs. It is possible for a system to have four phases: conceptual, definition, acquisition, and operational. The HQ USAF OPR for system program management will, through the system life cycle, be transferred to the next deputate having prime responsibility. Some of the major steps involved in most system programs are shown in Table 8. Key documents involved in the system life cycle are described in the following paragraphs.

1. <u>System Management Directives (SMD's)</u>

These directives provide uniform HQ USAF direction for initiating, changing, and terminating system programs under AFR 375-1. The first SMD establishes the charter for conducting a system program and will designate application of system management, transmit or reference the

TABLE 8 - HQ USAF SYSTEM PROGRAM RESPONSIBILITY

System Life Cycle	Deputy Chief of Staff OPR
Conceptual phase (concept formulation)	
Initial SMD (charter)	
$PTDP_1$ review--PCP processing	AFRDC (R&D) or AFSDC (S&L)
$PTDP_2$ review	
Memorandum or PCP processing	
Definition phase (contract definition)	
SMD issued	
PA issued	
Budget authority issued by AFABF (Director of Budget)	
FTA issued	AFRDC or AFSDC
Contractor selection	
Memorandum or PCP processing	
PSPP	
Acquisition phase	
SMD issued	
SPP review	
Contracting	
Development effort	
Production	
PCP/PA/BA	AFSDC
Category I, II tests	
Updating changes	
Last article delivered	
Transition agreement	
SMD issued	
Operational phase	AFXOP or other

current requirements document, and request a Program Change Proposal (PCP) and either a Preliminary Technical Development Plan (PTDP) or a Proposed System Package Plan (PSPP). If a formal definition phase is not planned, a PSPP is requested from the implementing command, not a PTDP. Although an SMD reflects policy decisions made within OSD and HQ USAF, including changes in the Force and Financial Plan (F&FP), an SMD in itself does not constitute authority to let a contract. An approved (signed) secretarial Determinations and Findings (D&F) is required before contract negotiations can be initiated or an RFP issued. Fund availability is established and a secretarial statement of Final Technical Approval (FTA) is obtained before a contract containing RDT&E funds may be signed. Separate program authorizations (PA's) issued by AFRRP (Assistant for R&D Programming) and Procurement Authorizations (PA's) issued by AFSPD provide procurement authorization.

2. <u>Program Change Proposal (PCP)</u>

This document, submitted by HQ USAF to the Secretary of Defense, introduces a new program to the F&FP or changes an approved program element in excess of established thresholds. A "proposed PCP" is submitted by AFSC to request an appropriate change to the program. The implementing command initially submits the PCP to the appropriate HQ USAF OPR along with a PTDP, PSPP, or other technical backup data attached.

3. <u>Preliminary Technical Development Plan (PTDP)</u>

This document is submitted by AFSC as the initial response to the RAD indicating approval of the ROC. The PTDP is used by HQ USAF to support the PCP submitted to OSD for approval of the definition phase.

4. <u>Proposed System Package Plan (PSPP)</u>

This document, normally prepared by AFSC, is submitted as a product of the definition phase or on direction of HQ USAF. It includes a system description, cost estimates, resource requirements, performance specifications, schedules, and related information for each alternative proposed. It should be definitive enough to allow incentive and/or fixed-price contracts to be negotiated in the acquisition phase.

5. <u>System Program Directive (SP Directive)</u>

This formal document, issued by HQ USAF, approves a system program defined in the PSPP and authorizes the publication of the SPP. The SP Directive identifies the availability of financial and other resources, the importance category, the impact on other Air Force programs, and other program direction. Subsequent program changes are made as amendments to the SP Directive.

6. <u>System Definition Directive (SDD)</u>

This is the formal document issued by HQ USAF approving the PTDP. The SDD identifies the availability of financial and other resources as applicable, provides authority to AFSC to establish a formal SPO, sets the parameters for the System Program Director (SPD), and establishes the roles of the participating organizations. The SDD also constitutes authority for solicitation of industry sources with the intent to commit the Government within approved fund authorizations.

7. <u>System Package Program (SPP)</u>

The SP Directive requires the System Program Director (SPD), who is head of the SPO and manager of the approved system program during the definition and acquisition phases, to convert the approved portions of the PSPP into the SPP. The SPP specifies the integrated and time-phased tasks and resources required of and by all participating organizations in acquiring and supporting the system.

A complete SPP consists of the following sections:

- Section 1: Program Summary
- Section 2: Schedules
- Section 3: Program Management
- Section 4: Intelligence Estimate
- Section 5: Operations
- Section 6: Acquisition
- Section 7: Civil Engineering
- Section 8: Logistics
- Section 9: Manpower and Organization
- Section 10: Personnel Training
- Section 11: Financial
- Section 12: Requirements
- Section 13: Authorizations
- Section 14: General Information
- Section 15: Security
- Section 16: Biomedical

In general, the Preliminary Technical Development Plan (PTDP) and the Proposed System Package Plan (PSPP) contain the same type of information and follow the same order. Section 14, General Information, must include (AFR 375-4) a description of all EDP systems used in support of the proposed system (but not an integral part of the system).

D. AFR 100 Series Regulations

The 100 series regulations deal, in general, with communications-electronics activities within the Air Force. In many instances, computers are involved in such systems; hence AFADA becomes involved in the approval cycle (AFR 300-2A).

AFR 100-2 defines a ground communications electronics meteorological (CEM) system as two or more physically separated but interdependent and interrelated equipment or facilities, complete with supporting structures and services. Ground CEM requirements can be of two types: quantitative and qualitative. A quantitative requirement is defined as a need for specific equipment or capability to accomplish a mission wherein the equipment or capability is available without further research and development effort. A qualitative requirement is defined as a need for a particular capability to accomplish a mission wherein the equipment or techniques must be researched or developed.

A qualitative ground CEM requirement is prepared and submitted to HQ USAF (AFORQ) as an ROC (Required Operational Capability). (AFR 57-3 previously required a QOR, but this regulation has been superseded by AFR 57-1, 17 June 1966.) After HQ USAF recognizes and validates a requirement, including OSD approval, presumably an RAD is issued. This document should describe the characteristics of the required CEM equipment and levy the requirement on AFSC to develop a new item of equipment or determine other means of satisfying the requirement. Implementation will be under AFM 100-18 or 375 series as directed by HQ USAF.

Quantitative ground CEM requirements are submitted to HQ USAF (AFSME) for validation as an Advance Communications-Electronic Requirements Plan (ACERP) or a Communications-Electronics Implementation Plan (CEIP). If data processing is involved, ACERP's and CEIP's are also submitted to AFADA and are accepted by this organization in lieu of DAP's.

The ACERP is a statement of a current or future need for ground CEM equipment or facilities that are available without further development or research. Approval of an ACERP by HQ USAF constitutes acknowledgement and recognition of the stated operational requirement (approval in principle) and authorizes preparing and processing a CEIP. In certain instances, the ACERP is accepted, CEIP requirements are waived, and AFLC is directed to implement the approved ACERP.

The CEIP is a detailed plan that provides information essential for final operational evaluation and programming actions.

E. <u>AFR 80-2, Documents Used in the Management of Air Force Research and Development</u>

AFR 300-7, Data Automation, R&D Support, specifically excludes ADP equipment developed for a particular use through expenditure of RDT&E funds. It is therefore possible for computing equipment to be acquired through submission of a development plan, as described in AFR 80-2, Attachment 2. Section 9c of these instructions requires only a minimum of data regarding EDP equipment.

APPENDIX G

SUMMARY OF CURRENT REPORTS COVERING AIR FORCE ADP EXPERIENCE AND ASSETS

Number	Name	Preparing Agency	Frequency	Contents
1 AF-E6	Data Systems Automation Program (Volume I)	HQ USAF (AFADAE)	Continually updated	Contains narrative description of all approved system design projects and active data systems. Includes a general description of files, records, input, output, and tie-in with other systems in addition to naming the organizations involved.
1 AF-E6	Data Systems Automation Program (Volume II)	HQ USAF (AFADAE)	Quarterly	Lists, by installation, all equipment systems installed or programmed for installation. Data include location, command, organization level, equipment type, procurement status, dates of installation, and planned phase-out, in addition to application title and implementation date.
4 AF-E6	Data Processing Equipment Inventory	Inventory is maintained by processing locations	Quarterly for leased, semi-annually for purchased; submitted to HQ USAF (AFADSCA)	Lists, by location, all data processing equipment used. Report includes location, machine or component serial number, acquisition code, and date of installation for each EDPE and PLAM component and piece of peripheral equipment.
6 AF-E6	ADPE Cost and Utilization	Installations assigned ADPE	Monthly to HQ USAF (AFADSCA)	Reports cost and utilization data for all data processing equipment assigned to the reporting installation. Data include location, command, machine type, number of like machines, acquisition code along with total hours of use, total hours of unscheduled maintenance, and actual maintenance costs.
8 AF-E6	EDPS by Application and Hours of Use	Installations having EDPS	Monthly to HQ USAF (AFADSCA)	Section I of this report lists application status code, primary and secondary program language code, master file size at close of reporting month, and the total number of actions which were direct input for this application during the reporting month. Section II is a utilization

Number	Name	Preparing Agency	Frequency	Contents
				summary listing, by application, production and preparation times, productive and operational use time, program development and maintenance time, chargeable loss time, set-up and idle times, and scheduled and unscheduled maintenance times, along with a projected average monthly use for the next 6 months.
None	Contract End Item Detail Specification (Computer Program)	Contractor	As required	Computer program contract end items (CPCEI) are required for each computer program. Data contain date of checkout, storage allocation map, and data base configuration, including a detailed description of each file and table as well as storage location of each computer program. Also included is the data organization and detailed definition of the computer program.
DD-I&L(SA) 678	DOD ADPE Program Reporting System	Installations assigned ADPE or major command headquarters	Annual to HQ USAF (AFADAEA)	Report consists of six formats: (1) For contractor operations only; inventory of ADPE giving command, ADPE make and model and installation/removal dates. (2) For contractor operations only; lists ADPE utilization and application including hours of scheduled and unscheduled maintenance, hours of production, program development, and maintenance and set-up; also gives application status codes and program language codes. (3) Gives ADP personnel costs, actual and projected man-years, operating costs, and capital investment. (4) Gives installation address. (5) Reports man-years utilized and salary costs by data systems within each command. (6) Presents analysis of costs of ADP contractual services.

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Planning Research Corporation 1100 Glendon Avenue Los Angeles, California 90024	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP Not Applicable

3. REPORT TITLE
Phase II Final Report on Use of Air Force ADP Experience to Assist Air Force ADP Management Volume II. Phase II Activities

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Final report for period 16 February 1966 to 15 December 1966

5. AUTHOR(S) (Last name, first name, initial)
Gradwohl, Alan J.
Beckwith, George S.
Wong, Stanton H.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
15 December 1966	156	None

8a. CONTRACT OR GRANT NO. AF 19(628)-5988 b. PROJECT NO. 7990 c. None d. None	9a. ORIGINATOR'S REPORT NUMBER(S) PRC R-932, Volume II 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) None

10. AVAILABILITY/LIMITATION NOTICES
Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
None	Directorate of Planning & Technology, ESD AF Systems Command L. G. Hanscom Field, Bedford, Mass.

13. ABSTRACT This is Volume II of a three-volume final report that covers Phase II of a three-phase project on the Use of Air Force ADP Experience to Assist Air Force ADP Management. In Phase I, a feasible concept and preliminary approach to using experience was synthesized; in Phase II, the approach was refined, the concept was validated, and the potential use of experience was broadened; and in Phase III, the improved and expanded approach will be implemented Air Force-wide.

Volume I of the final report covers the following: the history of the project; conclusions of Phase II and recommendations for Phase III; and summaries of Phase II activities, the Phase III concept and plan, and the pilot version of the ADP Experience Handbook and Primer. Volume II reviews the four major activities of Phase II: data collection, data analysis, ADP Experience Handbook development, and Phase III planning. Volume III presents the detailed Phase III operational concept and development plan, followed by a summary of cost and benefits.

This is Volume II, in which the four major activities of Phase II are described. The design of the data collection questionnaire was based on the ADPS model (a concept of a "total" ADPS) and the workload model representing attributes of an ADPS. Data were collected on a stratified 18-ADPS sample, and the statistical analysis of these data produced five cost estimation equations. In addition, the data were used to produce a seven-page system description of each ADPS, which became the core of the ADP Experience Handbook. A Phase III operational concept and development plan was also synthesized.

14. KEY WORDS	LINK A ROLE	LINK A WT	LINK B ROLE	LINK B WT	LINK C ROLE	LINK C WT
1. Data Processing Systems						
2. Classification						
3. Abstracts						
4. Measurement						
5. Analysis						
6. Effectiveness						
7. Mathematical Models						
8. Information Retrieval						
9. Indexes						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2*a*. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7*a*. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8*a*. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through ________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through ________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through ________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.